



A Reinforcement Learning Approach of Data Forwarding in Vehicular Networks

Pengfei Zhu[✉], Lejian Liao, and Xin Li

School of Computer Science, Beijing Institute of Technology, Beijing, China
zhu_pengfei0408@163.com, {liao1j,xinli}@bit.edu.cn

Abstract. As the basis of vehicle ad hoc networks, the method of forwarding data is one of the most important parts which ensures the stability and efficiency of network communication. However, the high-speed mobile vehicle nodes cause frequent changes of network topology and disconnections of network links, casting a big challenge to the performance of network data delivery. Data forwarding methods based on the prior knowledge of vehicle's trajectory are difficult to adapt to the changing vehicle trajectory in real world applications, while getting destination vehicles' positions in broadcast way are extremely costly. To solve the above problems, we have proposed an association state based optimized data forwarding method (ASODF) with the assistance of low loaded road side units (RSU). The proposed method maps the urban road network into a directed graph, utilizes the carry-forward mechanism and decomposes the data transmission into decision-making data forwarding at intersections and data delivery on roads. The vehicles carried data combine the destination nodes locations obtained by low loaded road side units and their locations into association states, and the association state optimization problem is formalized as a Reinforcement Learning problem with Markov Decision Process (MDP). We utilized the value iteration scheme to figure out the delay-optimal policy, which is further used to forward data packets to obtain the best delay of data transmission. Experiments based on a real vehicle trajectory data set demonstrate the effectiveness of our model ASODF.

1 Introduction

With the improvement of the vehicle information technology and the development of the wireless communications technology, Vehicular Ad-hoc Network (VANET) has developed rapidly. VANET is a self-organizing network which is specially designed for inter-vehicle communication based on Mobile Ad-hoc Network (MANET). The communication in VANET can be divided into three types, including vehicle to vehicle (V2V), vehicle to infrastructure (V2I) and infrastructure to vehicle (I2V). The core of Vehicular Network is the speed and efficiency and the security of data transmission in the network which consist of vehicles and road side units. However, the characteristics of large amount of nodes and quickly movement of nodes make it difficult to use the effective routing

protocol algorithms in Internet and wireless sensor networks, e.g., literature [1] proposed secure cell relay routing protocol, and also including some exists key management schemes in wireless sensor network [2], e.g, Du et al. proposed some key management schemes [3,4]. And some other exists work for the problems of wireless sensor network elaborated in literature [5] and the time synchronization scheme proposed in literature [6] also can't be applied in VANETs. Additionally, network performance is critical to the data transmission mechanism, due to that the performance of network are influenced by many factors, including instability and uncertainty of the channel quality. Therefore, some particular methods have been proposed to solve these problems.

The exists work of data transmission in Vehicular Network can be divided into three types. First is methods based on topology. This type of methods can also be divided into two subtype: Proactive and Reactive. **Destination Sequenced Distance Vector** (DSDV) [7] and **Optimized Link State Routing protocol** (OLSR) [8] are two classical Proactive methods. These type of methods have a obvious disadvantages is that the node needs to update the routing information at any time, which consumes a lot of bandwidth. **Dynamic Source Routing** (DSR) [9] and **Ad hoc on Demand Distance Vector routing** (AODV) are two reactive methods. But due to the use of broadcast mode, DSR has a poor scalability which is not suitable for such a large-scale mobile network ad hoc networks. AODV can suitable for large-scale network, but it still has some problem, e.g., larger network overhead and expired routing problem.

Second, since VANET has the characteristic of frequent changes in its topology, it is very difficult for network nodes to set up and maintain a stable routing table. Therefore, topology-based data transmission schemes are not suitable for vehicular networking. With the popularization of GPS equipment, the data transmission method based on geographic location has been proposed. **Greedy Perimeter Stateless Routing** (GPSR) [10] and **Geographic source routing** (GSR) [11] are two methods based on geographic location. GPSR uses a greedy model to transmission data which has a locally optimal problem. GSR is different to these model whose nodes are able to randomly move, it utilizes the fact that vehicle can only drive on road. So data transmission can only occurs at a intersection. The GSR does not take into account the real-time traffic conditions in the road network and may result in a lack of connectivity due to too few vehicles on the road sections selected at the intersection.

Third, due to the frequent disconnection of network links in vehicular ad hoc networks and the inability to establish end-to-end routing of source nodes to destination nodes, researchers creatively introduced the mechanism of tolerates time-delay networks and opportunistic networks into vehicular network, proposed a data transmission scheme based on store-and-forward and carry-forward mechanism. **Static Node Assisted Adaptive Routing** (SADV) [12] and **Vehicle Assisted Data Delivery** (VADD) [13] are two methods based on Store and Forward mechanism. SADV deploys static nodes at intersections, i.e., roadside units (RSU), to aid in the transmission of data. SADV draws on the VADD's section delay model and the optimal path selection. SADV utilizes the store-and-forward and carry-forward mechanism, which make it be a efficient

data routing solution. However, The need to deploy infrastructure at each intersection makes SADV unsuitable for large-scale network environments. VADD is a method that proposed for sparse environment. First, VADD extracts a delay model from real vehicle trajectory data. Then VADD calculates the total delay of packets from the current intersection to the destination node through the adjacent crossroads according to the delay model. Finally VADD ranks the total delay to select the optimal routing.

Fourth, the model based on store-and-forward and carry-forward mechanism is a good solution to the problem of link disconnection caused by sparse vehicles. But they are still do not take the road restrictions and human behavior patterns caused by a certain trajectory of the vehicle into consideration. Therefore, the models based on vehicle trajectory are proposed. These models can be divided into two types. One is the models in which the vehicles' trajectory is fixed in advance. Another is models based on trajectory prediction. **Anchor based Street and Traffic Aware Routing (A-STAR)** [14], **Geographical Opportunistic Routing (GeOpps)** [15] and **Mobile Gateway based Forwarding (MGF)** [16] are several models with fixed trajectory. The vehicle nodes in A-STAR model will choose the route with high connectivity, which will cause too heavy or even congestion. GeOpps can obtain the fixed trajectory of the vehicle node through the navigation system, and utilize the trajectory information to send packets to vehicles close to the destination node selectively. But due to too much dependency of trajectory, it is limited to the navigation system and driver's driving habits. MGF only use bus to transmission data which makes it only available on buses. Different to the above model, **Trajectory-Based Data Forwarding (TBD)** [17], **Trajectory-based Statistical Forwarding (TSF)** [18], **Shared-Trajectory-based Data Forwarding Scheme (STDFS)** [19], **Trajectory Improves Data Delivery in Vehicular Networks (Trajectory)** [20] and **Delay-Optimal Data Forwarding (OVDF)** [21] are several models with trajectory prediction. However, TBD and TSF only suitable for some certain situations. STDFS is not very reliable due to overdependence on the trajectory. OVDF also assists data transmission with the aid of bus fixed tracks. The Trajectory model adopt Markov Chain to do trajectory prediction, it is a efficient model.

Last, the models based on road side units (RSU) are another type of scheme of data transmission in vehicular network, including the above models of TBD, TSF, MGF, SADV, STDFS, OVDF. ROAMER [22] further used RSU to transmit data with the dependence of wired backbone network which is contrary to the concept of VANET transmit data from vehicle node, and this model has high requirements to RSU.

In summary, there are two problems: (1) data forwarding models based on the prior knowledge of vehicle trajectory assumptions are difficult to adapt to changing vehicle trajectories in real-world applications, (2) whereas broadcast network based approaches has a large network overhead when obtaining destination vehicle's positions. Inspired by the exists work, we proposed an association state based optimal data forwarding model (ASODF) to solve the above problem. Our model is a mixed model which include V2I and V2V data transmission.

2 Background

In this section, we give a brief review of Markov Decision Process and its value based methods.

2.1 Markov Decision Process

Markov Decision Process (MDP) is an optimal decision process based on the Markov process theory for stochastic dynamical systems. It is widely used to solve the sequential problems that need to make the best decisions at all stages [23]. A sequential decision problem with known environment dynamics is usually formalized as a MDP, which is characterized by a 5-tuple $\langle S, A, T, R, \gamma \rangle$, where S is the set of states and is non-empty, A is the set of actions and is also non-empty, $T : S \times A \rightarrow \Pi(S)$ is the transition function, it gives the probability of the next state when an agent execute a action $a \in A$ at state $s \in S$, where $\Pi(S)$ represent the set of probability distribution on S , R represent the Reward function, it give the immediate reward when an agent execute a action $a_t \in A$ at state $s_t \in S$ and the state transit to state $s_{t+1} \in S$, then the reward is $R_t = R(s_t, a_t, s_{t+1})$, γ is the discount factor to calculate the expected reward. MDP based on Markov Property, i.e., no post-efficiency. That means that the transition function $T(s_t, a_t, s_{t+1})$ is depends only upon the present state, whereas has no relate with past other state, i.e., $T(s_t, a_t, s_{t+1}) = P(s_{t+1}|s_t, a_t)$. The goal to solve a MDP question is obtain a optimal policy π , which gives the best decision of all state when an agent is making a decision, so that the agent can get most rewards finally.

When the original state is s_0 of an agent¹, agent will select and execute an action a_0 , then the environment will transit to next state s_1 , and agent will select and execute an action again, until it arrive terminal state. And the model will gives a optimal policy π when it converged through iteration. The policy give the optimal action of a state, i.e., $a = \pi(s)$.

Value function is always used to evaluate a policy. Value function also been called cumulative discount rewards, it gives a estimation of an agent will get finally from current state s_t , i.e.,

$$V^\pi(s_t) = E^\pi[R(s_t) + \gamma R(s_{t+1}) + \gamma^2 R(s_t + 2) + \dots] \tag{1}$$

We can easily transform it to a simple form according to Bellman Equation,

$$V^\pi(s_t) = R(s_t, a_t, s_{t+1}) + \gamma \sum p(s_{t+1}|s_t, a_t) V^\pi(s_{t+1}) \tag{2}$$

The optimal policy should be the policy which can gives the decisions to get most cumulative reward from each state. So, the most reward of each state s_i is:

$$V^*(s_i) = R(s_i, a_i, s_j) + \max_\pi \sum_{j \in S} p(s_j|s_i, a_i) V^*(s_j) \tag{3}$$

¹ MDP assume that agent can get true state of environment, i.e., $s^{agent} = s^{env}$.

So the optimal policy from state s_i to terminal state is:

$$\pi^*(s_i) = \arg \max_{a_i} \sum p(s_j|s_i, a_i)V(s_j) \tag{4}$$

The methods of solve MDP to get a optimal solution including value iteration and policy iteration and other linear programming methods. In this paper, we will use value iteration methods to solve MDP problem. The process of Value Iteration shows in literature [24].

2.2 Data Delivery on Road

Carry-forward mechanism over is a good way to overcome the shortcomings of frequent disconnection of VANET links, which are widely used in data transmission research of vehicular network. The carry-forward mechanism of data delivery model on road is shown in Fig. 1. When there are vehicles in the communication of the vehicle carried data and the vehicle is closer to next intersection than current vehicle, then select and delivery these data to the vehicle closest to next intersection. If no vehicle, then the vehicle will continue carry the data. This is a greedy model, i.e., select the best vehicle of current situation, so that data package can be delivery to next intersection with fastest speed and the minimum number of forward. Since the transmission process is composed with vehicle store and wireless forward, the delay of data transmission is affected by two factors, one is the vehicle density, and another is vehicle wireless device communication range. Learn from literature [13], we use ρ_{ij} represent the density of road e_{ij} , R represent the radius of wireless communication range. VADD assume that the distribution of distance of two vehicle meet the exponential distribution with parameter $1/\rho_{ij}$. So, the delay d_{ij} on the road e_{ij} is:

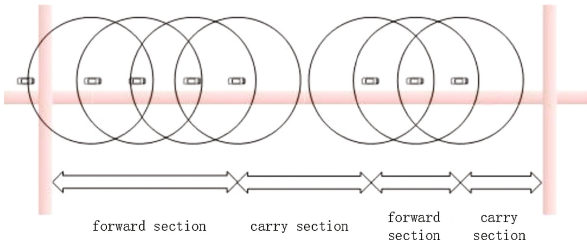


Fig. 1. Data delivery on road

$$d_{ij} = (1 - e^{-R \cdot \rho_{ij}}) \frac{l_{ij}c}{R} + e^{-R \cdot \rho_{ij}} \frac{l_{ij}}{v_{ij}} \tag{5}$$

Where l_{ij} represent the distance of road e_{ij} , c indicate the time needed to delivery data to next hop, v_{ij} represent the average speed on the road e_{ij} . These parameters can be obtained through the use of GPS devices by road traffic statistics or analysis of historical trajectory data.

2.3 Association State of Tag Game

Different to the exists application of MDP in data transmission, the state in MDP used in our paper is association state learn from Tag Game [25]. There are two roles in Tag Game, robot and opponent, The process is that the robot keeps chasing opponent until the robot catches up with opponent, which is shown in Fig. 2.

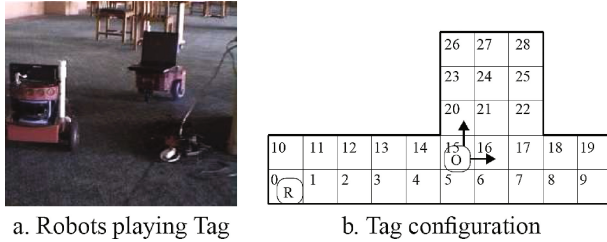


Fig. 2. Tag Game

Tag Game can be see as a Partially Observable MDP task, in which state are composed of the positions of robot and opponent, i.e., association state. The set of state of robot is $\{s_0, \dots, s_{29}\}$, the set of state of opponent is $\{s_0, \dots, s_{29}, s_{tagged}\}$, the association state is $s = \{Robot, Opponent\}$. The robot will execute one action of the set *North, South, West, East, Tag*, and then robot will get a immediate reward. When robot and opponent are in the same box, i.e., $Opponent = s_{tagged}$, then robot will catch up opponent and get the highest reward and game over.

3 Association State Based Data Forwarding Model

MDP has widely used to solve sequential decision-making tasks. Since the vehicle carrying data will meet other vehicles with different probabilities, the data forward in VANET can be formed to a Sequential decision making problem. We can consider the process of transmitting a data from source node to destination vehicle as the robot chasing the opponent in Tag Game, i.e., the data package chasing the destination vehicle and the vehicle carrying data are keep changing. In this section, we will form this problem to a MDP task.

3.1 Association State

Association State is the core of getting the position of destination vehicle dynamically and forwarding data optimally. With the use of association state, we can add the position information of destination vehicle to the MDP model and optimize the network delay of data transmission dynamically.

In our model, we use the current intersection as the current vehicle’s state. So the association state consists of the intersection of source vehicle and the intersection of destination vehicle. When the source vehicle node obtains the intersection information of the destination vehicle node, the low-load roadside unit and its wired backbone-assisted communication are utilized. We assume that each vehicle will register information on the roadside unit when it enter the coverage area of roadside unit of a intersection. Once the roadside unit detects the destination vehicle node, the intersection information will be transmitted to the roadside unit closest to the source vehicle node through the roadside unit backbone network. Then, the position information of destination vehicle node will be transmitted to the source vehicle node so that it can get the current association state. Due to the use of the roadside unit backbone network, the time delay can be ignored. The next state is also a necessary condition for solving a MDP problem.

In our model, the information of destination vehicle including speed, position, direction and etc. We can also obtain the next intersection of the current vehicle when it has not enter the coverage of the next intersection yet through the this information. Therefore, we can get the next association state.

3.2 Decision-Making of Association State

As we has described the transmission on road section, we will show the process of data transmission at a intersection. We will give priorities on the directions which is shown in Fig. 3 according to a fixed policy like VADD [13], where 1 represent the best direction to transmit data, 2 represent the second optimal direction, etc. We will select the optimal direction, i.e., priority is 1. It will transmission data when there is a vehicle in that direction, or will check if it is driving in this direction, if so then don’t forward to another vehicle, if not then will select vehicle in the second optimal direction, and etc.

At intersection i , decisions (actions) can be formed as a vector set $U(i)$, where $\pi_i^1 \pi_i^2 \pi_i^3 \dots \pi_i^{m_i} \in E$ is all m_i road sections connected with intersection i and the order indicates the priority. Our goals is to select the best decision (action) from set $U(i)$ to transmit data at current intersection.

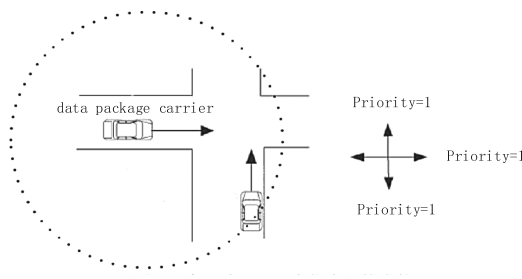


Fig. 3. The decision making of a intersection

3.3 Transition Probability

We use $P(s, \pi_i, s')$ represent the transition probability, where state s consist of the intersections of source vehicle and destination vehicle. So $P(s, \pi_i, s')$ is consist of two parts, the transition probability of source vehicle $P(src_t, \pi_i, src_{t+1})$ and the transition probability of destination vehicle turn to the next intersection $P(dest_t, dest_{t+1})$.

Assume that the current intersection is i and the policy is π_i , $P_{ij}(\pi_i) = P(src_t, \pi_i, src_{t+1})$ represent the probability that the data be transmitted to next intersection j along the road e_{ij} where $src_t = i, src_{t+1} = j$.

(1) *Computation of $P(src_t, \pi_i, src_{t+1})$.* We define three probability events:

- **A** represent the event that a vehicle has not met a vehicle which is heading to a road section with a higher priority than road segment e_{ij} .
- **B** represent the event that a vehicle met a vehicle which is drive to the road e_{ij} at intersection i and the vehicle itself doesn't drive to a road whose direction has a higher priority that e_{ij} .
- **C** represent the event that a vehicle drive to the road e_{ij} .

With the above definition, we can derive the probability of $P_{ij}(\pi_i)$:

$$\begin{aligned}
 p(src_t, \pi_i, src_{t+1}) &= P_{ij}(\pi_i) \\
 &= P[A \cap (B \cup C)] \\
 &= P(A) \times P(B \cup C) \\
 &= P(A) \times [P(B) + P(C) - P(B|C)P(C)] \\
 &= [\prod_{e_{ik} \in HPe_{ij}(\pi_i)} (1 - p_{ik})] \\
 &\quad \times [p_{ij} \times (1 - \sum_{e_{ik} \in HPe_{ij}(\pi_i)} p'_{ik}) + p'_{ij} - p_{ij} \times p'_{ij}]
 \end{aligned} \tag{6}$$

where $P(A)$ indicate the probability that event **A** occurred, $HPe_{ij}(\pi_i)$ represent the set of roads that have a higher priority that road e_{ij} . p_{ij} represent the probability that a vehicle drive from intersection i to intersection j , p'_{ij} represent the probability that a vehicle meet other vehicles which drive from intersection i to road e_{ij} . In our model, we set $p_{ij} = \frac{\#num(i \rightarrow j)}{\#num(i)}$, where $\#num(i \rightarrow j)$ is the number of vehicles drive to intersection j when it is at intersection i , $\#num(i)$ is the number of all vehicles reach intersection i . And we set $p'_{ij} = \frac{\#num^{met}(i \rightarrow j)}{\#num^{met}(i)}$, where $\#num^{met}(i \rightarrow j)$ is the number of vehicles that the vehicle met at intersection i and drive to intersection j , $\#num^{met}(i)$ is the number of all vehicles that the vehicle met at intersection i .

(2) *Computation of $P(dest_t, dest_{t+1})$.* In our model, we set $P(dest_t, dest_{t+1}) = \frac{\#num(dest_t \rightarrow dest_{t+1})}{\#num(dest_t)}$, where $\#num(dest_t \rightarrow dest_{t+1})$ is the number of vehicles that reach intersection i at time t and reach intersection j at time $t + 1$, and $\#num(dest_t)$ is the number of all vehicles reach intersection i at time t .

So the complete association state transition probability $P(s, \pi_i, s')$ of a carrier vehicle reached intersection i at time t is:

$$\begin{aligned}
 P(s, \pi_i, s') &= P(s_{t+1} = s | a_t = \pi, s_t = s') \\
 &= P((src, des)_{t+1} | \pi_i, (src, des)_t) \\
 &= P(src_t, \pi_i, src_{t+1}) * P(des_t, des_{t+1}) \\
 &= P_{ij}(\pi_i) * P(des_t, des_t + 1)
 \end{aligned}
 \tag{7}$$

3.4 Model Derivation

Network time delay is an important indicator of VANET performance. Its value is the accumulation of time delays on the roads that the data package passed, which is corresponds to the composition of the value function in MDP. So we use the time delay as MDP reward.

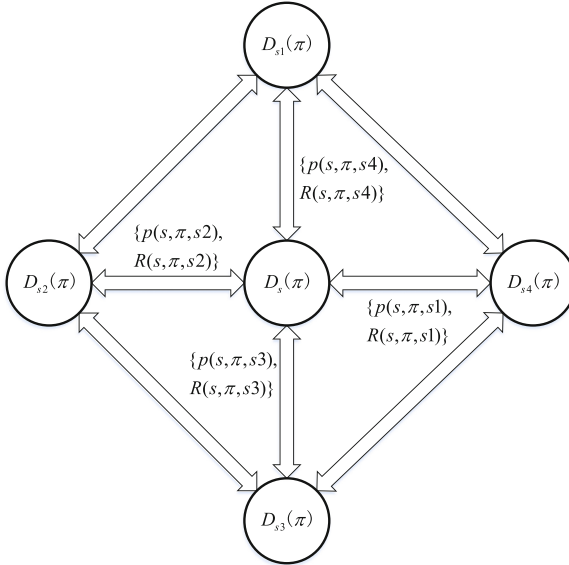


Fig. 4. The Markov Decision Process at state s

Assume that there are four adjacent state at state s , then the transition model from state s is shown in Fig. 4. The $D_s(\pi)$ represent the value function from state s , i.e., the estimated total time delay from the source vehicle which is at state s . So $D_s(\pi)$ can be formed as:

$$D_s(\pi_s) = \sum_{s' \in N(s)} P(s, \pi_s, s') \times [R(s, \pi_s, s') + D_{s'}(\pi)]
 \tag{8}$$

$$\begin{aligned}
&= P_{s,s_1} \times (R(s, \pi_s, s_1) + D_{s_1}(\pi_s)) \\
&+ P_{s,s_2} \times (R(s, \pi_s, s_2) + D_{s_2}(\pi_s)) \\
&+ P_{s,s_3} \times (R(s, \pi_s, s_3) + D_{s_3}(\pi_s)) \\
&+ P_{s,s_4} \times (R(s, \pi_s, s_4) + D_{s_4}(\pi_s))
\end{aligned} \tag{9}$$

The our goal is to minimize the total time delay, i.e.,

$$\min_{\pi} D_s(\pi), \forall s \tag{10}$$

And the optimal policy we will get is:

$$\pi^* = \langle \pi_s^*, \forall s \in S \rangle \tag{11}$$

The reward $R(s, \pi_i, s')$ we will get is derived as:

$$R(s, \pi_i, s') = R((src_t, des_t), \pi_i, (src_{t+1}, des_{t+1})) \tag{12}$$

$$= \frac{1}{2}(d_{src_{t+1}, src_t} + d_{des_{t+1}, des_t}) \tag{13}$$

3.5 Algorithm

Since our model is still a standard MDP model, we can use the standard Value Iteration to solve this question. The Algorithm is shown in Algorithm 1.

Algorithm 1. ASODF:Association State based Optimal Data Forwarding model

Input: Initial the Values of all state to D_0 , max iteration τ and threshold θ .

Output: The optimal policy $\pi^* = \langle \pi_s^*, \forall s \in S \rangle$ and the corresponding expected time delay $D_s^*(\pi^*)$

1: initial $g_0 = 0$ and $d_0 = 0$; **Local:** $k = 0$

2: **repeat**

3: $D_s^{k+1} = \sum_{s' \in N(s)} T(s, \pi_s, s') \times (d_{s,s'} + D_{s'}^{k+1})$

4: $\pi_s^{k+1} = \arg \min_{\pi_s \in \Pi(s)} \sum_{s' \in N(s)} T(s, \pi_s, s') \times (d_{s,s'} + D_{s'}^k)$

5: $k = k + 1$

6: **until** $\max_{s \in S} \|D_s^k - D_s^{k-1}\| < \theta$ or $k > \tau$

7: $\pi_s^* = \pi_s^k, \pi^* = \langle \pi_s^*, \forall s \in S \rangle$

8: **return** $\pi^*, D_s^*(\pi^*)$

4 Experiments

In this section, we will describe our experiments in detail.

4.1 DataSet

In order to make the experimental results more real and convincing, we run experiments on a real vehicle data set of SUVnet of Shanghai [26]. It include 5000 taxis and buses' trajectories data, and we only use the part of taxis' data.

We preprocessed the data set, including:

- Clean the data, including remove duplicate data and error data.
- Repair drift data based on road structure.
- Since the taxi data recorded an average of every 30s, we interpolated the discontinuous trajectory data and error trajectory data.

4.2 Experiment Settings

We select about 2700 taxis as our object vehicles. In our experiments, we randomly select 200 vehicles as our the source vehicle and the destination vehicles to transmit data. We assume that each data package is the same size. And some hyper parameters are shown in Table 1.

Table 1. Parameter setting of experiments

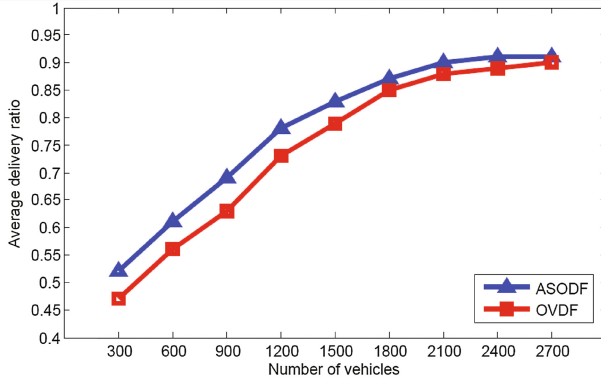
Parameters	Values (range)
Wireless transmission range	200 m
The number of experimental vehicles	300 to 2700
θ	0.001
τ	1000
The number of vehicles carried data	200
Time difference to the next intersection	10 s
Time to live (TTL)	1 h, 2 h, 3 h

4.3 Experiment Result Analysis

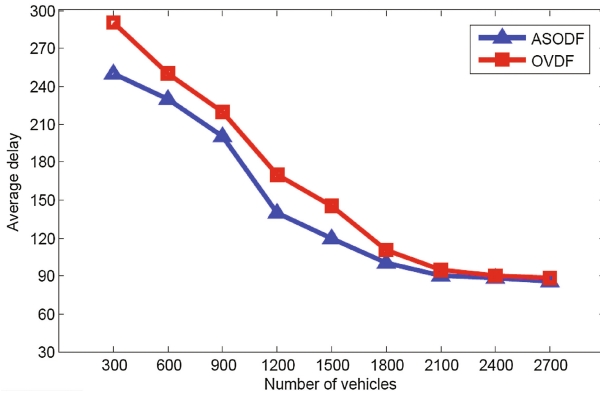
We compared our model ASODF with OVDF-P, which is one of models of OVDF. In OVDF, a data package is successfully be transmitted when it is be transmit to a road side unit and we changed this setting. In our experiments, a data is transmitted successfully when it is transmitted to a moving vehicles.

The results is of average delivery ratio and average delay is shown in Fig. 5(a) and (b) respectively.

The results shown in Fig. 5(a) demonstrate that our model has a high delivery ratio when the number of vehicles are the same, i.e., the density of vehicles is same. And a obvious conclusion is that both model's average delivery ratio will increase with the increase of the number of vehicles. Fig. 5(b) shows that our model has a low delay when the number of vehicles is small, i.e, our model are better when the vehicles in network is sparse. and both the model will have a similar results with the increase of the vehicles.



(a) Results of average delivery ratio



(b) Results of average delay

Fig. 5. POMDP to MDP generalization performance

Table 2. Comparison of average delivery ratio

Vehicle number	ASODF (s)	OVDF (s)	Improvement
300	0.52	0.47	10.64%
600	0.61	0.56	8.93%
900	0.69	0.63	9.52%
1200	0.78	0.73	6.85%
1500	0.83	0.79	5.06%
1800	0.87	0.85	2.35%
2100	0.90	0.88	2.27%
2400	0.91	0.89	2.25%
2700	0.92	0.90	2.22%

Table 3. Comparison of average delay

Vehicle number	ASODF	OVDF	Improvement
300	250	290	13.79%
600	230	250	8.71%
900	200	220	9.09%
1200	140	170	17.65%
1500	120	145	17.24%
1800	100	110	9.10%
2100	90	95	5.26%
2400	88	90	2.52%
2700	86	89	3.37%

The result of comparing two models is shown in Tables 2 and 3. Table 2 shows that with the increase of vehicles, the improvement of our model compare to OVDF tend to a small value of 2.22%. The reason is that with the increase of vehicle density, fewer and fewer vehicle communication links are disconnected, and more and more data packets are transmitted by wireless, the improvement is gradually reduced. And Table 3 shows that the average delay improvement is 13.79% at a low vehicle density. In summary, that our model are better that OVDF, particularly at a low vehicle density.

5 Conclusion

In this paper, we proposed an association state based optimal data forwarding model (ASODF) to improve the data delivery ratio and decrease the delivery delay in VANET. Our model formed the data forwarding to a reinforcement learning tasks and use standard value iteration method to solve it. And Experiments show that our model can get a high delivery ratio and a lower delay, particularly, our model can do better in deal with sparse environment in VANET.

References

1. Du, X., Xiao, Y., Chen, H.-H., Wu, Q.: Secure cell relay routing protocol for sensor networks. *Wirel. Commun. Mob. Comput.* **6**(3), 375–391 (2006)
2. Xiao, Y., Rayi, V.K., Sun, B., Du, X., Hu, F., Galloway, M.: A survey of key management schemes in wireless sensor networks. *Comput. Commun.* **30**(11), 2314–2341 (2007)
3. Du, X., Xiao, Y., Guizani, M., Chen, H.-H.: An effective key management scheme for heterogeneous sensor networks. *Ad Hoc Netw.* **5**(1), 24–34 (2007)
4. Du, X., Guizani, M., Xiao, Y., Chen, H.: Transactions papers a routing-driven elliptic curve cryptography based key management scheme for heterogeneous sensor networks. *IEEE Trans. Wirel. Commun.* **8**(3), 1223–1229 (2009). <https://doi.org/10.1109/TWC.2009.060598>

5. Du, X., Chen, H.-H.: Security in wireless sensor networks. *IEEE Wirel. Commun.* **15**(4) (2008)
6. Du, X., Guizani, M., Xiao, Y., Chen, H.-H.: Secure and efficient time synchronization in heterogeneous sensor networks. *IEEE Trans. Veh. Technol.* **57**(4), 2387–2394 (2008)
7. Perkins, C.E., Bhagwat, P.: Highly dynamic destination-sequenced distance-vector routing (DSDV) for mobile computers. *ACM SIGCOMM Comput. Commun. Rev.* **24**(4), 234–244 (1994)
8. Clausen, T., Jacquet, P.: Optimized link state routing protocol (OLSR). Technical report (2003)
9. Lee, S.-J., Gerla, M., Chiang, C.-C.: The dynamic source routing protocol for multi-hop wireless adhoc networks
10. Karp, B., Kung, H.-T.: GPSR: greedy perimeter stateless routing for wireless networks. In: *Proceedings of the 6th Annual International Conference on Mobile Computing and Networking*, pp. 243–254. ACM (2000)
11. Lochert, C., Hartenstein, H., Tian, J., Fussler, H., Hermann, D., Mauve, M.: A routing strategy for vehicular ad hoc networks in city environments. In: *Proceedings of the Intelligent Vehicles Symposium*, pp. 156–161. IEEE (2003)
12. Ding, Y., Xiao, L.: SADV: static-node-assisted adaptive data dissemination in vehicular networks. *IEEE Trans. Veh. Technol.* **59**(5), 2445–2455 (2010)
13. Zhao, J., Cao, G.: VADD: vehicle-assisted data delivery in vehicular ad hoc networks. *IEEE Trans. Veh. Technol.* **57**(3), 1910–1922 (2008)
14. Costa, P., Frey, D., Migliavacca, M., Mottola, L.: Towards lightweight information dissemination in inter-vehicular networks. In: *Proceedings of the 3rd International Workshop on Vehicular Ad Hoc Networks*, pp. 20–29. ACM (2006)
15. Leontiadis, I., Mascolo, C.: GEOPPS: geographical opportunistic routing for vehicular networks. In: *IEEE International Symposium on World of Wireless, Mobile and Multimedia Networks: WoWMoM 2007*, pp. 1–6. IEEE (2007)
16. Chen, L., Li, Z.-J., Jiang, S.-X., Feng, C.: MGF: mobile gateway based forwarding for infrastructure-to-vehicle data delivery in vehicular ad hoc networks. *Jisuanji Xuebao (Chin. J. Comput.)* **35**(3), 454–463 (2012)
17. Jeong, J., Guo, S., Gu, Y., He, T., Du, D. TBD: trajectory-based data forwarding for light-traffic vehicular networks. In: *29th IEEE International Conference on Distributed Computing Systems, ICDCS 2009*, pp. 231–238. IEEE (2009)
18. Jeong, J., Guo, S., Gu, Y., He, T., Du, D.H.: TSF: trajectory-based statistical forwarding for infrastructure-to-vehicle data delivery in vehicular networks. In: *2010 IEEE 30th International Conference on Distributed Computing Systems (ICDCS)*, pp. 557–566. IEEE (2010)
19. Xu, F., Guo, S., Jeong, J., Gu, Y., Cao, Q., Liu, M., He, T.: Utilizing shared vehicle trajectories for data forwarding in vehicular networks. In: *2011 Proceedings of IEEE INFOCOM*, pp. 441–445. IEEE (2011)
20. Wu, Y., Zhu, Y., Li, B.: Trajectory improves data delivery in vehicular networks. In: *2011 Proceedings of IEEE INFOCOM*, pp. 2183–2191. IEEE (2011)
21. Choi, O., Kim, S., Jeong, J., Lee, H.-W., Chong, S.: Delay-optimal data forwarding in vehicular sensor networks. *IEEE Trans. Veh. Technol.* **65**(8), 6389–6402 (2016)
22. Mershad, K., Artail, H.: Performance analysis of routing in VANETs using the RSU network. In: *2011 IEEE 7th International Conference on Wireless and Mobile Computing, Networking and Communications (WiMob)*, pp. 89–96. IEEE (2011)
23. Bellman, R.: A Markovian decision process. *J. Math. Mech.* **6**, 679–684 (1957)
24. Sutton, R.S., Barto, A.G., *Reinforcement Learning: An Introduction*, vol. 1, no. 1. MIT Press, Cambridge (1998)

25. Pineau, J., Gordon, G., Thrun, S., et al.: Point-based value iteration: an anytime algorithm for POMDPs. In: IJCAI, vol. 3, pp. 1025–1032 (2003)
26. Huang, H.-Y., Luo, P.-E., Li, M., Li, D., Li, X., Shu, W., Wu, M.-Y.: Performance evaluation of SUVnet with real-time traffic data. *IEEE Trans. Veh. Technol.* **56**(6), 3381–3396 (2007)