

Research Article

A Fuzzy-Rule Based Data Delivery Scheme in VANETs with Intelligent Speed Prediction and Relay Selection

Yi Zhou ^{1,2}, Huanhuan Li,^{1,2} Chenhao Shi,^{1,2} Ning Lu,³ and Nan Cheng⁴

¹School of Computer and Information Engineering, Henan University, Kaifeng 475004, China

²International Joint Research Laboratory for Cooperative Vehicular Networks of Henan, Henan, China

³Department of Computing Science, Thompson Rivers University, Kamloops, BC V2C 0C8, Canada

⁴Department of Electrical and Computer Engineering, University of Toronto, Toronto, ON M5S 1A1, Canada

Correspondence should be addressed to Yi Zhou; zhouyi@henu.edu.cn

Received 15 December 2017; Revised 29 May 2018; Accepted 9 July 2018; Published 8 August 2018

Academic Editor: Enrico Natalizio

Copyright © 2018 Yi Zhou et al. This is an open access article distributed under the Creative Commons Attribution License, which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.

Data delivery in vehicular networks (VANETs) is a challenging task due to the high mobility and constant topological changes. In common routing protocols, multihop V2V communications suffer from higher network delay and lower packet delivery ratio (PDR), and excessive dependence on GPS may pose threat on individual privacy. In this paper, we propose a novel data delivery scheme for vehicular networks in urban environments, which can improve the routing performance without relying on GPS. A fuzzy-rule-based wireless transmission approach is designed to optimize the relay selection considering multiple factors comprehensively, including vehicle speed, driving direction, hop count, and connection time. Wireless V2V transmission and wired transmissions among RSUs are both utilized, since wired transmissions can reduce the delay and improve the reliability. Each RSU is equipped with a machine learning system (MLS) to make the selected relay link more reliably without GPS through predicting vehicle speed at next moment. Experiments show the validity and rationality of the proposed method.

1. Introduction

Vehicular Network (VANET) is a special type of the Mobile ad-hoc network (MANET) where every single node is a vehicle moving on the road. In addition to safety and privacy concerns [1], the challenges of studying VANET are mainly from the special characteristics of such networks: frequently link disconnections, rapidly topology changing, and large-scale sizes.

There are many important applications of VANETs, which are related to safety and nonsafety [2, 3]. Almost all applications are inseparable from message transmission; that is, data delivery is the cornerstone for the wider deployment of VANET applications [4]. In VANETs, the network protocols might fail due to frequent link disruptions caused by various factors, such as severe interference, interceptions, radio channel fading, and frequent topology changes. As a result, the connections are intermittent and many network services fail to function properly and their performance

is seriously degraded. Therefore, developing efficient data delivery schemes with the presence of link disruptions is of great practical importance [5].

Fuzzy logic, known for its ability to deal with complex and imprecise problems, is a very promising technology in such a dynamic and complex context [6]. Meanwhile, massive traffic data in the transport system have important guiding significance to data delivery [7]. To efficiently utilize the dense data, algorithms skilled in dealing with historical data such as machine learning (ML) are needed. In order to improve the data transmission quality (DTQ), we propose a novel message delivery scheme which optimizes the V2V communication by optimizing the selection of relay nodes with the help of fuzzy rules and avoids the usage of GPS with the assistance of machine learning algorithms. For the unstable communication links in wireless transmission, we design a fuzzy-rule-based approach to select an optimal path from all the possible paths with considering multiple factors comprehensively. To reduce the usage of V2V

communication, RSUs are installed in the scene, which will decrease the network delay and packets loss probability by means of the rapid and stable transmission in a wired network. To deal with the dynamic of VANET, every RSU is embedded with a KNN-based machine learning system (MLS) to provide estimations about movements of vehicles and travelling paths.

The contributions of the paper are threefold, (1) our fuzzy-rule-based wireless transmission method can determine the optimal relay node in V2V communication, (2) a novel vehicle-based short-term speed prediction method with high practicability and flexibility is designed to enhance link stability, and (3) to defend client privacy and ensure system scalability our KNN-based machine learning system embedded in RSU supports GPS-free dynamic vehicle location prediction. The remainder of this paper is organized as follows. Section 2 describes the related work. Section 3 introduces the proposed model. Section 4 gives the detailed analysis of our system. Section 5 presents related simulation results. Section 6 concludes this paper.

2. Related Work

Xiang X et al. proposed two self-adaptive on-demand geographic routing protocols [8]. By adopting different schemes, the two protocols can obtain and maintain local topology information on data traffic demand. SOGR-HR (SOGR with Hybrid Reactive Mechanism) purely relies on one-hop topology information for forwarding as other geographic routing schemes; SOGR-GR (SOGR with Geographic Reactive Mechanism) combines both geographic and topology-based mechanisms for more efficient path building. The proposed SOGR-GR protocol achieves a better balance between control overhead and packet forwarding overhead. However, the core mechanism of SOGR-GR is purely based on traffic conditions and demands without considering the nodes' dynamicity and density. Our proposed strategy intelligently captures the dynamic changes in traffic operation through our machine learning system embedded in each RSU.

In [9], a machine learning assisted route selection (MARS) system is proposed to design routing protocols for urban environment. In order to predict the moves of vehicles and choose some suitable routing paths with better transmission capacity, the widely applied machine learning algorithm K-means is adopted. As an unsupervised cluster approach, K-means will judge the similarity of the data and decide which data can be grouped into the same cluster. However, sometimes this algorithm may be trapped into local optimum and different initial condition may produce different result because it is sensitive to initial centroids. Moreover, arithmetic mean is not robust to outliers, and very far data from the centroid may pull the centroid away from the real one. In our proposal, to avoid falling into local optimum, we select KNN algorithm instead of K-means.

In [10], a novel Fuzzy Logic based Greedy Routing (FLGR) protocol which focuses on transmitting safety messages with minimum delay is designed. The proposed FLGR is a position based greedy routing protocol that uses multiple metrics of neighbor vehicles to decide which vehicle is the

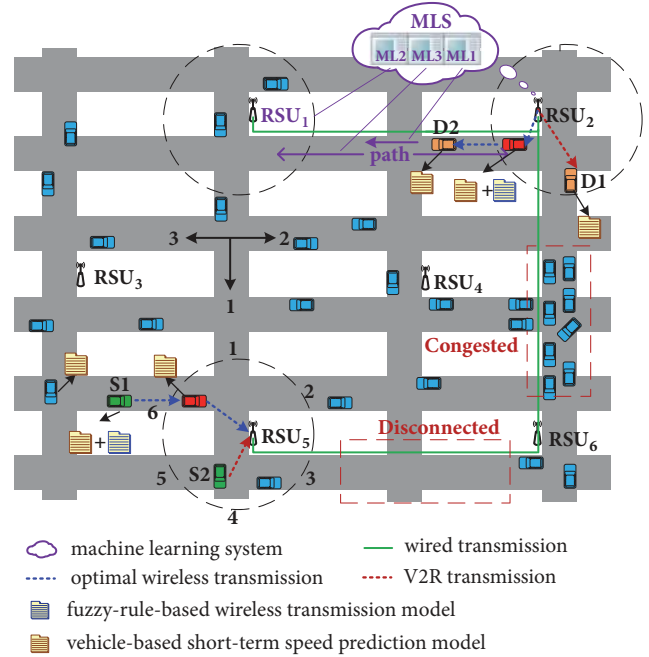


FIGURE 1: Data delivery model.

best relay node by employing fuzzy logic. It selects the node with maximum distance, speed, and progress and minimum angular deviation from current forwarding node towards destination as the next hop. However, this protocol only considered the current state of vehicles and did not take the impact of future state into account when making decisions. In our fuzzy-rule-based method, we design a novel vehicle-based short-term speed prediction method to take future speed of vehicles into consideration.

3. Proposed Model

In traditional routing protocols, V2V communication is the primary transmission mode. However, the vehicle density can greatly affect such protocols: (a) if the vehicle density is too high, competition for channel resources may incur collisions and result in lower packet delivery rate (PDR), and (b) if the vehicle density is too low, carry-and-forward technique will increase transmission delay in road segments with poor communication connectivity. In case the road segment is congested or disconnected, as shown in Figure 1, DTQ will be degraded. In addition, the selection of relay node plays the decisive role in the effect of wireless transmission. Not all the vehicles in transmission range are qualified for the relay node. A bad choice not only fails to deliver packet accurately and rapidly but also leads to high delay even packet loss. In general, the vehicle with the same driving direction as the source vehicle and having similar travelling speed with the source vehicle is more promising to be a good choice. Of course, other factors should also be taken into account if we intend to find the optimal choice. In order to connect to a certain vehicle, GPS is leveraged in most routing protocols to locate that node. However, excessive dependence

on GPS is unreliable: (a) GPS information is not always available especially in shielded areas, (b) GPS may induce some security issues and invade personal privacy, and (c) GPS sensor is power utilization equipment that sometimes users may turn off to decrease the power consumption, e.g., when a mobile phone is used as GPS device [11].

To solve the problems mentioned above, we propose a novel message delivery scheme which optimizes the V2V communication with the help of RSUs and fuzzy rules, and avoids the usage of GPS with the assistance of machine learning algorithms. Wired links between RSUs via backbone network can decrease the use of V2V communication obviously so as to reduce wireless transmission, so that not only higher packet delivery ratio but also lower network delay time are guaranteed. For the unavoidable wireless transmission portion, we design a fuzzy-rule-based method which takes driving direction, vehicle speed, connection time, and hop count into consideration to make the relay nodes selection more reasonable. To enhance the stability of V2V transmission link, a novel vehicle-based short-term vehicle speed prediction method is proposed to predict vehicle speed dynamically. In our proposed scheme, every RSU is embedded with a specially designed machine learning system to process dynamic traffic information and provide routing decisions. Our GPS-free dynamic vehicle location prediction can acquire the location of destination vehicle to eliminate those potential troubles caused by GPS.

We use the following notations:

- (i) S: the source node
- (ii) D: the destination node
- (iii) RSU_s : the RSU which the source vehicle is connecting to
- (iv) RSU_d : the RSU which the destination vehicle is connecting to or has just left
- (v) RSU_n : the RSU which the destination will visit next
- (vi) Blind zone: areas not covered by any RS

In the process of delivering packets, the source vehicle will first attempt to access a RSU. If source vehicle is covered by RSU_s , V2R communication will be selected, since On Board Unit (OBU) can communicate with RSU employing Dedicated Short Range Communication (DSRC) technology to implement information transmission [12, 13]. If source node is in the blind zone, the fuzzy-rule-based approach will be employed for establishing the communication link. First, it will find potential paths among all the routes. On the basis of fuzzy mathematics, four factors (driving direction, vehicle speed, connection time, and hop count) are considered. During this process, a novel vehicle-based short-term vehicle speed prediction method is proposed to predict vehicle speed dynamically. And then, evaluate every potential path. Each path will be evaluated according to the fuzzy comprehensive evaluation method (FCEM) and assigned an integrated assessment value and its corresponding evaluation grade based on the maximum membership principle. Finally, select wireless transmission path according to the evaluation results. Our scheme will first select the optimal path to

delivery packets; in case the RSU_s fails to receive the packets after a tolerable threshold time, the suboptimum path will be selected.

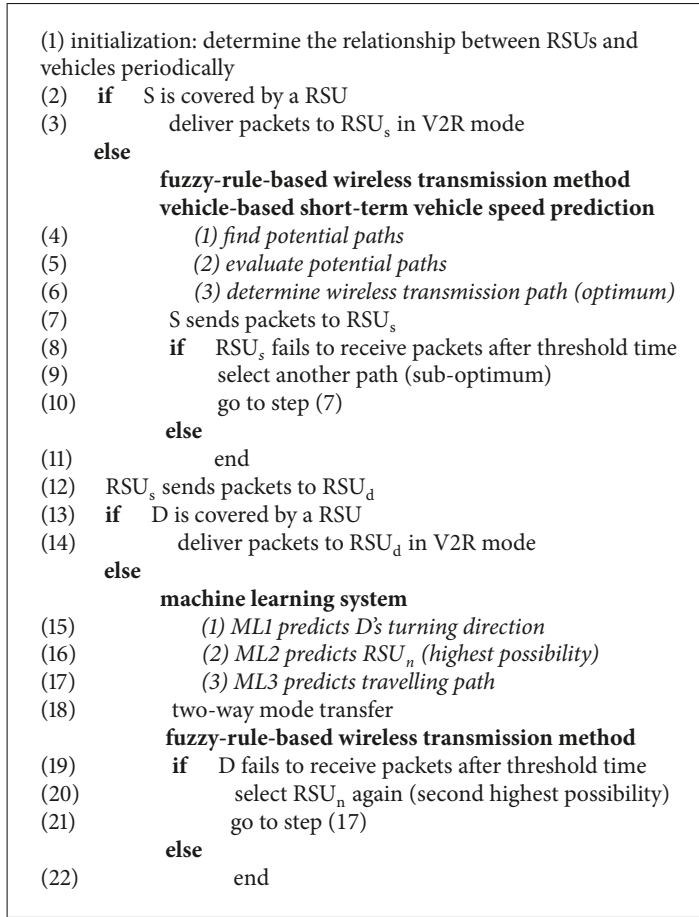
After receiving the packets, RSU_s will send the Transmission Request (TREQ) to other RSUs to find RSU_d . Each RSU will maintain a Node Table, which lists the vehicles in its coverage in this moment and those vehicles before with the time when they left. Every Node Table stores the traffic information collected by RSU in real time and is updated periodically. By checking respective tables, each RSU can judge whether to respond to the TREQ or not. Only RSU_d sends the Transmission Reply (TREP) to RSU_s to inform the location state of the destination vehicle, and then RSU_s delivers packets to RSU_d . RSUs installed on the road sides are important components for the Intelligent Transportation System (ITS). Backbone network makes it possible to interconnect between RSUs which is used between RSU_s and RSU_d . Compared with wireless networks, wired networks are more stable and speedy, especially in such a highly dynamic environment.

Finally, the packets will be delivered to the D in V2R mode in case it is covered by a RSU, otherwise the transmission path will be predicted with the assist of machine learning system. First, machine learning 1 (ML1) will predict destination's turning direction at next intersection after leaving RSU_d : go straight, turn left, or turn right. And then, machine learning 2 (ML2) will predict the probability that destination node travels into each RSU to determine RSU_n . Last, machine learning 3 (ML3) will predict the travelling path from a certain exit of RSU_d to RSU_n . After the travelling path is determined, packets will be transferred in a two-way mode; that is, both RSU_d and RSU_n are dedicate to searching connected ways to D along the predicted path. In this section, fuzzy-rule-based approach is used again. If RSU_d finds the path faster than RSU_n , it will deliver packets to D in unicast mode and, meanwhile, inform RSU_n to stop searching. Otherwise, RSU_d will relay packets to RSU_n by wired way to complete the dissemination. In case the destination vehicle fails to receive packets after a tolerable threshold time, the scheme will select the RSU with the second highest possibility and perform the two-way transfer again and so on.

As shown in Figure 1, S2 can transmit packets to RSU_5 via wireless V2R communication, while S1 needs fuzzy-rule-based approach for the selection of a relay vehicle. And then after checking Node Table, RSU_5 delivers packets to RSU_2 by means of wired transmission. For destination vehicle D1, RSU_2 can deliver packets to it directly. For destination vehicle D2, a transmission path will be established with the assistance of machine learning system in RSU_2 . Our proposal makes data transmission less dependent on vehicle density and avoids the negative issues resulted from the usage of GPS. Algorithm 1 presents the whole message delivery process concisely. More detailed information will be described in next part.

4. Data Delivery Scheme

4.1. Fuzzy-Rule-Based Wireless Transmission Method. Multihop broadcasting schemes are particularly preferred in



ALGORITHM 1

wireless transmission to transmit information to vehicles or infrastructures that cannot communicate directly in VANET. However, ordinary broadcasting may suffer from frequent contention and serious collision [14] and thus cause broadcast storms [15]. How to suppress broadcast storm is one of the important issues in the wireless transmission process [16, 17]. In order to alleviate broadcast storm and guarantee fast and efficient messages dissemination, this paper proposes a fuzzy-rule-based wireless transmission approach to provide an optimal forwarder selection scheme.

4.1.1. Structure of Wireless Transmission Method

(a) *Finding Potential Paths.* First, the source vehicle will initiate a Routing Request (RREQ) in its transmission range, which indicates whether or not it can reach a RSU. Then neighboring nodes, hearing this request advertisement, will rebroadcast this beacon to their neighbors. Finally RSU receiving this beacon will send back a Routing Reply (RREP) to announce the RSU_s and the route. This process is repeated until any one of the terminal conditions is met, (a) preset end time and (b) preset the number of routes that satisfy the following conditions. All the potential routes found should guarantee that the RREQ is heard by some RSU within a certain number of hops (default two hops). As shown in

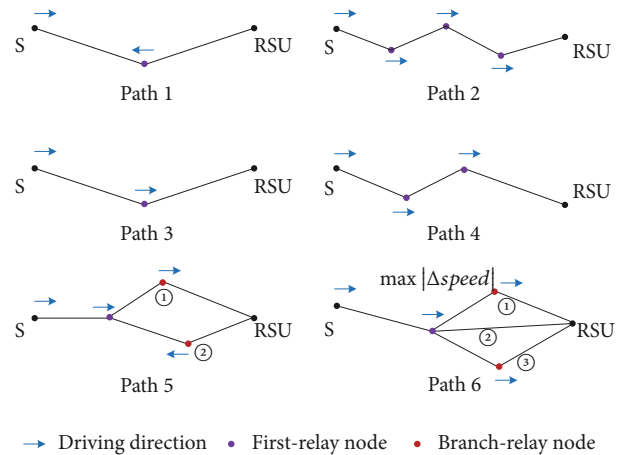


FIGURE 2: Judgment of potential paths.

Figure 2, all the found paths are qualified as potential routes except for path1 and path2. If there are two or more branching paths after the first-relay node, pre-evaluation is needed to select branch-relay node. The principle of pre-evaluation is defined as follows: (1) abandon branch-relay nodes driving in the opposite direction or with the maximum speed difference

with the first-relay node; (2) the subpath without branch-relay node is deemed to have the same direction and speed with the first-relay node; (3) the upper bound of the number of subpath is two. As shown in Figure 2, after pre-evaluation, path5 will select ① as the branch-relay node and path6 will select ② and ③, and path3 and path4 do not need pre-evaluation.

(b) *Evaluating Potential Paths.* Paths are evaluated according to the fuzzy comprehensive evaluation method, wherein the index weight vector is determined based on the analytic hierarchy process (AHP). Fuzzy logic, first mentioned by Zadeh in 1965 [18], allows uncertain information to be processed by using simple IF-THEN rules. This laid the foundation for the future development of fuzzy theory [19]. Known for its ability to deal with complexity and imprecision problems, fuzzy logic is a powerful mathematical tool to deal with multiparameter problems in such a dynamic and complex context [20]. On the basis of fuzzy mathematics, four factors are considered to derive a fuzzy relation matrix R for each potential path. And according to this matrix, our method will assign every potential route an integrated assessment value and its corresponding evaluation grade.

(c) *Determining Wireless Transmission Path.* The last stage of our fuzzy-ruled-based wireless transmission approach is to select wireless transmission path according to the evaluation results. Our scheme will first select the optimal path to delivery packets; in case the RSU fails to receive the packets after a tolerable threshold time, the suboptimum path will be selected. For paths belonging to different evaluation grades, higher grades have priority over lower grades; for paths belonging to the same grade, higher assessment value has priority over lower assessment value.

For example, in Figure 3, vehicle s sends out a RREQ and its neighbors will rebroadcast this beacon. As such, through vehicle a and vehicle b as the relay nodes, a RSU can be found to receive the RREQ and then reply to RREP. In this way, we find a potential path $s \rightarrow a \rightarrow b \rightarrow RSU$. After route evaluation, the optimal path for the wireless communication connection to a RSU will be determined. Now, source node can unicast the packets to that RSU along the path carried by RREP.

4.1.2. *Relay Node Selection.* With a view to select a more reliability relay node, we employ FCEM to combine several influential factors to conduct an evaluation on each potential path. FCEM is a scientific assessment method based on fuzzy mathematics with the considerations of a plurality of influence factors [21]. It uses fuzzy logic to systematically perform evaluation of real world systems that are not clearly defined.

(a) *Confirming the Evaluation Index Set.* In order to evaluate whether or not a vehicle is qualified for the next relay node, index set is defined as $U = \{f_1, f_2, f_3, f_4\} = \{|\Delta V_n|, |\Delta D|, HC, CT\}$, where f_1 is the absolute value of speed difference at next moment between current vehicle and the potential vehicle, f_2 is the absolute value of driving

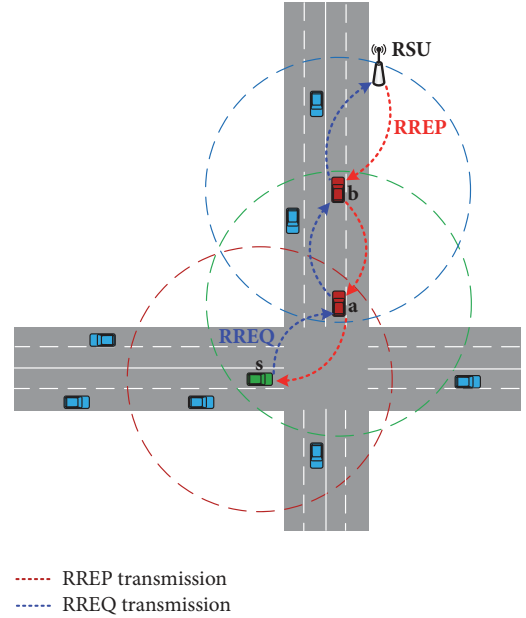


FIGURE 3: Finding potential wireless transmission path.

direction difference between that two vehicles, f_3 is the hop count from the potential vehicle to a RSU, and f_4 is the corresponding connection time. The speed difference is the first index to measure a potential vehicle since similar speed is a necessary condition for stable vehicle distance. The concrete description of vehicle speed prediction will be presented in the next section. The driving direction difference is another key index, if two vehicles are driving in the opposite direction and the stability of the link must be hard to ensure. In addition, hop count and connection time are another two important influencing factors, and too many hops or too long connection time will affect the link stability seriously.

(b) *Confirming the Evaluation Criteria Set.* The degree of satisfaction is measured using the so-called remark set that consists of a set of linguistic variables such as “good” or “bad” [22]. In order to evaluate whether or not a vehicle is qualified for the next relay node, we define the evaluation criteria set with four ratings $v = \{very\ good, good, general, bad\}$.

(c) *Determining the Index Weight Vector.* According to theory of the analytic hierarchy process, the index weight vector is set to be $P = [0.32\ 0.4492\ 0.0957\ 0.1351]$. More detailed information about the AHP method will be introduced in the following.

(d) *Constructing the Fuzzy Relation Matrix.* The fuzzy relation matrix is defined as (1), where r_{ij} indicates the membership of the i_{th} index belonging to the j_{th} rate. The membership function is established according to the characteristics of the index system. For the discrete variables f_2 and f_3 , membership grade is determined according to Tables 1 and 2. For continuous variables f_1 and f_4 , the membership function is defined as shown in Figures 4 and 5. For variable f_4 , fuzzy set function is calculated as (2)–(5), and we can get the

TABLE 1: Membership function of f_2 .

f_2	v			
	v1	v2	v3	v4
0	1	0	0	0
1	0	0.5	0.5	0
2	0	0	0	1

TABLE 2: Membership function of f_3 .

f_3	v			
	v1	v2	v3	v4
0	1	0	0	0
1	0	1	0	0
2	0	0	1	0

membership function of variable f_1 by substituting variate $v/100$ for t .

$$R = \begin{bmatrix} \mu_{v1}^{f_1} & \mu_{v2}^{f_1} & \mu_{v3}^{f_1} & \mu_{v4}^{f_1} \\ \mu_{v1}^{f_2} & \mu_{v2}^{f_2} & \mu_{v3}^{f_2} & \mu_{v4}^{f_2} \\ \mu_{v1}^{f_3} & \mu_{v2}^{f_3} & \mu_{v3}^{f_3} & \mu_{v4}^{f_3} \\ \mu_{v1}^{f_4} & \mu_{v2}^{f_4} & \mu_{v3}^{f_4} & \mu_{v4}^{f_4} \end{bmatrix} = (r_{ij})_{m \times n} \quad (1)$$

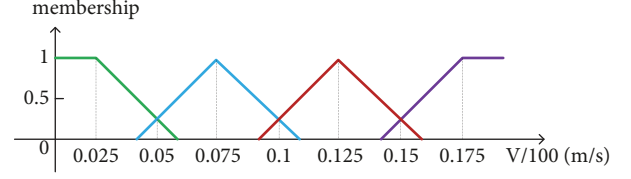
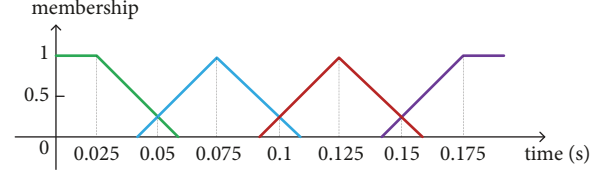
$$\mu_{v1}^{f_4} = \begin{cases} 1 & 0 \leq t \leq \frac{1}{40} \\ -30t + 1.75 & \frac{1}{40} \leq t \leq \frac{7}{120} \\ 0 & \text{others} \end{cases} \quad (2)$$

$$\mu_{v2}^{f_4} = \begin{cases} 30t - 1.25 & \frac{1}{24} \leq t \leq \frac{3}{40} \\ -30t + 3.25 & \frac{3}{40} \leq t \leq \frac{13}{120} \\ 0 & \text{others} \end{cases} \quad (3)$$

$$\mu_{v3}^{f_4} = \begin{cases} 30t - 2.75 & \frac{11}{120} \leq t \leq \frac{1}{8} \\ -30t + 4.75 & \frac{1}{8} \leq t \leq \frac{19}{120} \\ 0 & \text{others} \end{cases} \quad (4)$$

$$\mu_{v4}^{f_4} = \begin{cases} 30t - 4.25 & \frac{17}{120} \leq t \leq \frac{7}{40} \\ 1 & \frac{7}{40} \leq t \\ 0 & \text{others} \end{cases} \quad (5)$$

(e) *Determining Comprehensive Evaluation Class.* By performing the fuzzy composite operation between the index weight vector and the fuzzy relation matrix, a comprehensive evaluation vector model is established as shown in (6), where b_1, b_2, b_3 , and b_4 represent the four ranks of the evaluation

FIGURE 4: Membership function of f_1 .FIGURE 5: Membership function of f_4 .

set, respectively. We determine the corresponding evaluation grade according to the maximum membership principle.

$$B = P \circ R = [b_1, b_2, b_3, b_4] \quad (6)$$

4.1.3. *Index Weight Vector.* The index weight vector represents the different weights of the selected four impact factors in the forwarding node selection. In order to perform the fuzzy composite operation between the index weight vector P and the fuzzy relation matrix R , we need to determine values for vector P exactly. In this paper, we employ analytic hierarchy process [23] to assign the optimal weight for every index to identify optimal forwarder. AHP decomposes the complex problem into a hierarchy of subproblems to evaluate the relative importance of each criterion [24]. The alternatives are chosen according to their weights towards each criterion and ultimately towards the goal [25].

(a) *Analytic Hierarchy Structure.* An important part of AHP is to structure the analytic hierarchy: (i) to state the objective; (ii) to define the criteria; (iii) to choose the alternatives [15]. Based on requirements in the scenario stated in this paper, the objective (the top level) is to select an optimal forwarder. The criteria (the medium level) include the absolute value of speed difference between current vehicle and the potential vehicle, the absolute value of direction difference between two vehicles, the hop count from the potential vehicle to a RSU, and the corresponding connection time. The alternatives (the bottom level) include all the candidate nodes within the communication range of the source node. The hierarchical tree is formed based on these three layers as shown in Figure 6, where node1, ..., nodeC represent the candidate forwarder vehicles.

(b) *Judgment Matrix.* According to the degree of importance to relay node selection, the importance ranking of four influence factors is defined as $f_2 > f_1 > f_4 > f_3$. The criteria are pairwise compared to find their importance towards the goal, and such a pairwise comparison is represented as a judgment matrix shown in (7), where n is the total number

TABLE 3: Relative importance between criteria.

A_{ij}	definition
1	i and j are equally important
3	i is moderately more important than j
5	i is strongly more important than j
1/3	i is moderately less important than j
1/5	i is strongly less important than j
2,4, 1/2, 1/4	intermediate values

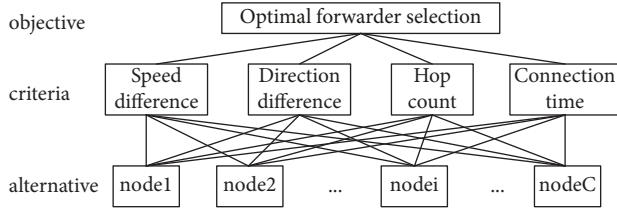


FIGURE 6: Analytic hierarchy structure.

of criteria. In the matrix, A_{ij} denotes the relative importance of criteria i to j . $A_{ij} = 1$ indicates that index i is as important as index j , $A_{ij} > 1$ indicates that index i is more important than index j , and $0 < A_{ij} < 1$ indicates that index i is less important than index j . The property of judgment matrix is also shown in (7) [25]. The relative importance of one criterion over another can be expressed in pairwise comparison matrix according to Table 3.

$$\begin{array}{c|cccc}
 A & f_1 & f_2 & f_3 & f_4 \\
 \hline
 f_1 & A_{11} & A_{12} & A_{13} & A_{14} \\
 f_2 & A_{21} & A_{22} & A_{23} & A_{24} \\
 f_3 & A_{31} & A_{32} & A_{33} & A_{34} \\
 f_4 & A_{41} & A_{42} & A_{43} & A_{44}
 \end{array} \quad (7)$$

$$(1) A_{ij} > 0, (2) A_{ij} = \frac{1}{A_{ji}} \quad (i \neq j), (3) A_{ij} = 1 \quad (i = j)$$

As such, we get the judgment matrix A, and matrix B is the result after normalization.

$$A = \begin{bmatrix} 1 & \frac{1}{2} & 4 & 3 \\ 2 & 1 & 4 & 3 \\ \frac{1}{4} & \frac{1}{3} & 1 & \frac{1}{2} \\ \frac{1}{3} & \frac{1}{4} & 2 & 1 \end{bmatrix} \quad (8)$$

$$B = \begin{bmatrix} 0.2857 & 0.2308 & 0.3636 & 0.4 \\ 0.5714 & 0.4615 & 0.3636 & 0.4 \\ 0.0714 & 0.1538 & 0.0909 & 0.0667 \\ 0.0714 & 0.1538 & 0.1818 & 0.1333 \end{bmatrix}$$

(c) *Consistency Examination.* The last step of AHP is consistency examination to check whether the comparison matrix

A is consistent or not [26]. The judgment errors are detected using the consistency ratio (CR), which is the ratio of the consistency index (CI) to the random index (RI). The CI value is calculated as (9), where n is the number of decision factors, λ_{\max} is maximal eigenvalue of matrix A, and RI values are shown in Table 4 [27]. The errors in judgments are considered tolerable when $CR \leq 0.1$; otherwise, the pairwise comparisons need to be adjusted [25]. After calculation, the consistency ratio $CR = 0.0526 < 0.1$ in the judgment matrix, so the matrix A meets the compliance requirements. After normalization, the index weight vector is set to be $P = [0.32 \ 0.4492 \ 0.0957 \ 0.1351]$.

$$CI = \frac{\lambda_{\max} - n}{n - 1} = \frac{4.1421 - 4}{4 - 1} = 0.0474 \quad (9)$$

$$CR = 0.0526 < 0.1$$

$$\lambda_{\max} = 4.1421 \quad (10)$$

$$CR = \frac{CI}{RI} < 0.1$$

4.2. Vehicle-Based Short-Term Speed Prediction. Short-term vehicle speed prediction is one of the most critical components of an ITS. Real-time and accurate vehicle speed prediction is the key to traffic control and traffic guidance and provides important information for intelligent vehicles and transportation applications.

Although there has been a growing body of studies on short-term vehicle speed prediction approaches, most methods belong to segment-based methods [28, 29]. Segment-based models predict vehicle speed for a certain road by analyzing historical traffic data collected from one or more road fracture surfaces. For example, a short-term traffic speed prediction model which predicts the traffic speed on a route containing more than one road link is developed based on a support vector machine model in [28]. Based on the SVM algorithm, the temporal information of the target road link and traffic speed of upstream/downstream road links are considered. Segment-based methods are suitable for vehicle navigation if historical traffic data about related road segments are acquired, but the inherent disadvantage of these methods is the poor scalability. The predictive ability is restricted by the vehicle location that is whether or not our database contains the historical traffic information about the road vehicle is travelling on. Besides, segment-based vehicle speed prediction model cannot capture the subtle fluctuations caused by routine traffic flows and the sudden disruption caused by accidents, since those abnormal data are often eliminated in data process stage.

In order to deal with these problems, a novel vehicle-based short-term vehicle speed prediction model, based on the weighted K-Nearest Neighbor algorithm (W-KNN), is introduced in this study. The predictive ability of our model is restricted by related traffic data which are obtained by OBU employing DSRC technology instead of the vehicle location. Since vehicle speed prediction is based on the latest data, the impact caused by subtle fluctuations will soon be reflected. The travel speed at next moment is affected by speed of the

TABLE 4: Mean random consistency index RI.

n	1	2	3	4	5	6	7	8	9	...
RI	0	0	0.58	0.9	1.12	1.24	1.32	1.41	1.45	...

moment and speed in the past, and the closer the time, the greater the impact. In addition, in an urban road network, road links do not exist in isolation. Traffic conditions on both upstream and downstream road segments can affect the vehicle speed of the current road segment [28]. To improve the prediction accuracy, we synthetically consider limited spatial and temporal influence factors in our proposal. For the temporal domain, we dynamically select real-time traffic data and historical data within certain time lags to provide valuable sample data for each prediction. For the spatial domain, the testing vehicle and vehicles in its communication range are considered since vehicle's travel speed is affected by other vehicles around.

Based on the analysis above, we determine the feature vector $[V, a, C, \Delta C]$, which is composed of velocity, acceleration, the vehicle count in the testing sample's communication range, and the vehicle count gradient. For each potential neighbor $x(i, j)$, the state vector is defined as

$$\begin{aligned} S(i, j) &= [V(i, j), a(i, j), C(i, j), \Delta C(i, j)] \\ &= [V_i(t - j\Delta t), a_i(t - j\Delta t), C_i(t - j\Delta t), \\ &\quad \Delta C_i(t - j\Delta t)] \quad j = 1, 2, \dots, l, \quad i = 0, 1, 2, \dots, C, \end{aligned} \quad (11)$$

where $i = 0$ indicates the testing vehicle, $i = 1, 2, \dots, C$ indicates other vehicles in its communication range, and C is the count of vehicles in its communication range. Δt is the time interval between two data collections. l is the number of time interval Δt to determine the time lag of the historical data. The label for potential neighbor $x(i, j)$ will be $y(i, j) = V_i(t - (j - 1)\Delta t)$. For example, a training data collected from the third vehicle in the testing vehicle's communication range with two time interval lags can be written as $S(3, 2) = [V_3(t - 2\Delta t), a_3(t - 2\Delta t), C_3(t - 2\Delta t), \Delta C_3(t - 2\Delta t)]$, and the corresponding label is $y(3, 2) = V_3(t - \Delta t)$. In order to predict a vehicle's travel speed at next moment, we should provide the input vector at time t $X = [V_0(t), a_0(t), C_0(t), \Delta C_0(t)]$ to the prediction model, and then predicted result $\widehat{V}_0(t + \Delta t)$ will be output for this testing vehicle $X_0(t)$.

We have analyzed and compared three different prediction models to select a suitable one with higher accuracy in this section. The moving average data-based (MAD) model adopts one of the simplest techniques, a straight average of the dependent variables. In this model, the travel speed at next moment is predicted by the speed in the previous time period. The data used for prediction is the previous data closest to the testing time t [28]. The observed equation of the moving average data-based model is shown in

$$\widehat{V}_0(t + \Delta t) = \frac{1}{1+n} \sum V_0(t - j\Delta t) \quad j = 0, 1, 2, \dots, n, \quad (12)$$

where n is the number of previous time periods used in model. This does consider all the dependent variables evenly

but does not consider the relationship between the testing data and each training data. The differences between samples are neglected, and each training sample is considered to make the same contribution. And also, this model does not take the impact from spatial domain into account. Several simulation experiments to explore the relationship between parameter n and the prediction accuracy are conducted as shown in Figure 2.

K-Nearest Neighbor learning, one of the most popular realizations of IBL (instance based learning), combines the target values of K selected neighbors to predict the target value of a given test pattern. Once the state vector is defined, the next step is the selection of a suitable determinant to measure the closeness between the testing sample and each candidate neighbor in the training data set. Rank results based on the closeness information will determine the member of the neighborhood. The similarity is usually based on the Minkowski distance metrics, wherein the L_r distance, as written in (13), is referred to as $r = \{1, 2, \dots, \max\}$ in the {Manhattan distance, Euclidean distance, ..., Chebyshev distance} metrics [29]. Due to the dynamic nature of VANETs, the traffic condition shows successive fluctuations. In other words, the time-series traffic state is a highly dynamical system with uncertain noise, which is a meaningful signal for the future state. The Euclidean distance is sensitive to noise, so the abnormal variation can be real-timely captured when the current state is either seriously disturbed or rapidly changed. And this is why Euclidean distance has been most frequently used to measure the similarity in NPR-based (non-parametric regression) traffic variable prediction [30–32].

$$L_r = \left[\sum |x(i, j) - X_0(t)|^r \right]^{(1/r)} \quad (13)$$

In this study, the Euclidean distance ($r = 2$), $ED(i, j)$, between $x(i, j)$ and $X_0(t)$, is used, which can be written as (14). Given a testing sample $V_0(t)$, the Euclidean distance metric is used to obtain the K -nearest neighbors and their corresponding labels from the training data set. Hence, the potential output vector Y_i , associated with the i th nearest neighbors, is made up of two values as shown in (15). The observed equation with nonweighted KNN model is shown in (16).

$$\begin{aligned} ED(i, j) &= d(x(i, j), X_0(t)) = \left[|V(i, j) - V_0(t)|^2 \right. \\ &\quad \left. + |a(i, j) - a_0(t)|^2 + |C(i, j) - C_0(t)|^2 \right. \\ &\quad \left. + |\Delta C(i, j) - \Delta C_0(t)|^2 \right]^{1/2} \end{aligned} \quad (14)$$

$$Y_i = [V_i, ED_i] \quad i = 1, 2, \dots, K \quad (15)$$

$$\widehat{V}_0(t + \Delta t) = \frac{1}{K} \sum V_i \quad i = 1, 2, \dots, K \quad (16)$$

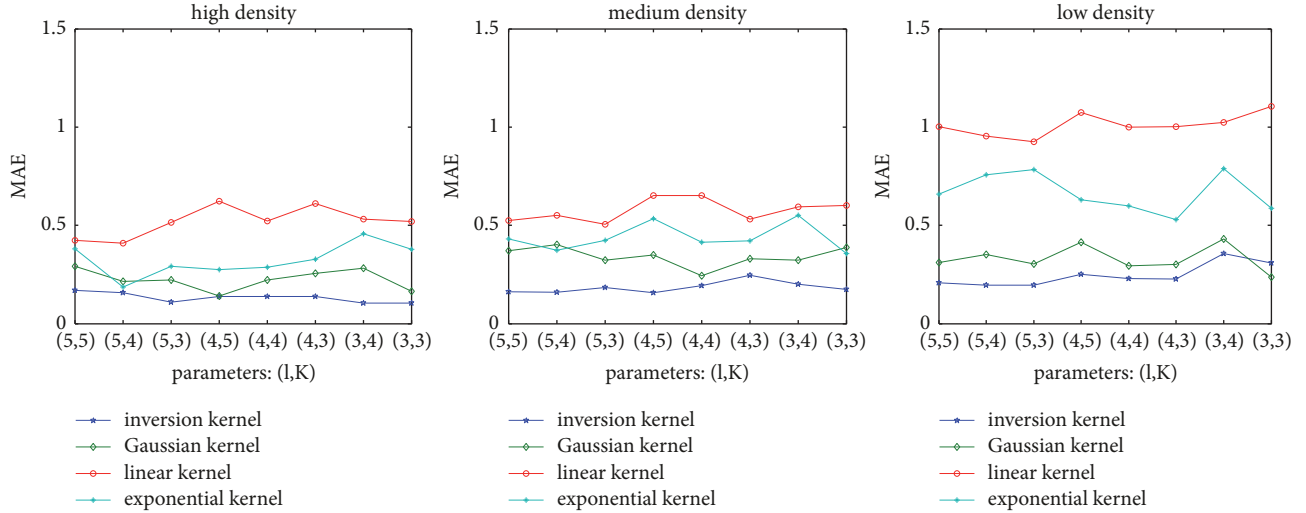


FIGURE 7: Results of three kernel functions.

One major challenging issue of the nonweighted KNN model is that its performance strongly depends on one key model parameter: the number of nearest neighbors K . In fact, no well-established method exists for selecting an optimal K when using the KNN algorithm. The number of nearest neighbors is often chosen empirically by cross-validation or domain experts in practice [33]. Therefore, a simple empirical or experimental test [3–5] is sufficient to find a suitable K -value. KNN considers the correlation between the testing vehicle and other vehicles by selecting the K -nearest neighbors according to the Euclidean distance metrics, but the K selected neighbors are treated equally without considering their differences according to (16).

For weighted-KNN model, in addition to the number of neighbors, the weights about those neighbors are another key parameter. Regarding the weights to the neighbors, the rule of thumb so far is “A father neighbor gets a smaller weight” [33]. That is a farther neighbor receives a smaller weight, which reduces its effect on the prediction results compared to other closer neighbors. There are a number of well-known kernel functions, which decrease monotonically as distance increases, such as the linear kernel $w_i = 1 - d(x_i, X_0(t))$ [34], the inversion kernel $w_i = (d(x_i, X_0(t)))^{-1}$ [35], the exponential kernel $w_i = \exp(-d(x_i, X_0(t)))$ [36], and the Gaussian kernel $w_i = \exp(-d(x_i, X_0(t))^2)$ [37]. Atkerson et al. [35] claimed that there is no clear evidence that any kernel function is always superior to the others, but some outperformed others on some data sets [38]. In order to select an appropriate kernel function to acquire better prediction precision for our weighted-KNN model, we performed simulation experiments using these four kernel functions. MAE (Mean Absolute Error) [24], as shown in (19), is introduced to evaluate each kernel. The larger the value of MAE, the greater the prediction error, and in contrary, the lower prediction accuracy. As shown in Figure 7, in terms of MAE, the inversion kernel which endows different weights for every element in the K select samples by the inverse of the

corresponding Euclidean distances (18) is the best choice in our model.

$$\widehat{V}_0(t + \Delta t) = \sum w_i V_i \quad i = 1, 2, \dots, K \quad (17)$$

$$w_i = \frac{d_i^{-1}}{\sum d_i^{-1}} \quad \text{s.t.} \sum_i w_i = 1 \quad (18)$$

$$MAE = \frac{1}{n} \sum |\widehat{V}_i - V_i| \quad i = 1, 2, \dots, n \quad (19)$$

To evaluate the prediction performances of the short-term traffic speed prediction models, we compare the moving average data-based model, the pure KNN model, and the weighted-KNN model utilizing the traffic data from SUMO microscopic traffic simulator. We simulate different traffic conditions by changing the number of vehicles: 150 vehicles for sparse status in Figure 8(a), 450 vehicles for normal status in Figure 8(b), and 750 vehicles for congested status in Figure 8(c). To optimize the prediction performance for each model, a series of contrastive experiments with different combinations of influential parameters are built in different traffic conditions. Results in Figure 8 show that our weighted-KNN model with spatial-temporal parameters exhibits better performance compared with the pure KNN model and the MAD model in terms of prediction accuracy. And for different time interval Δt , the MAD model acquires the worst MAE no matter the vehicle density is low, medium, or high. The results of the W-KNN model are better than or equal to that of the KNN model under all the various experimental conditions listed. Based on the experimental results, we adopted the W-KNN model to predict the short-term vehicle speed in this paper.

4.3. Machine Learning System. In our proposed mechanism, machine learning system is indispensable when the destination is in the blind zone. The main duty of our specially designed machine learning system embedded in every RSU is

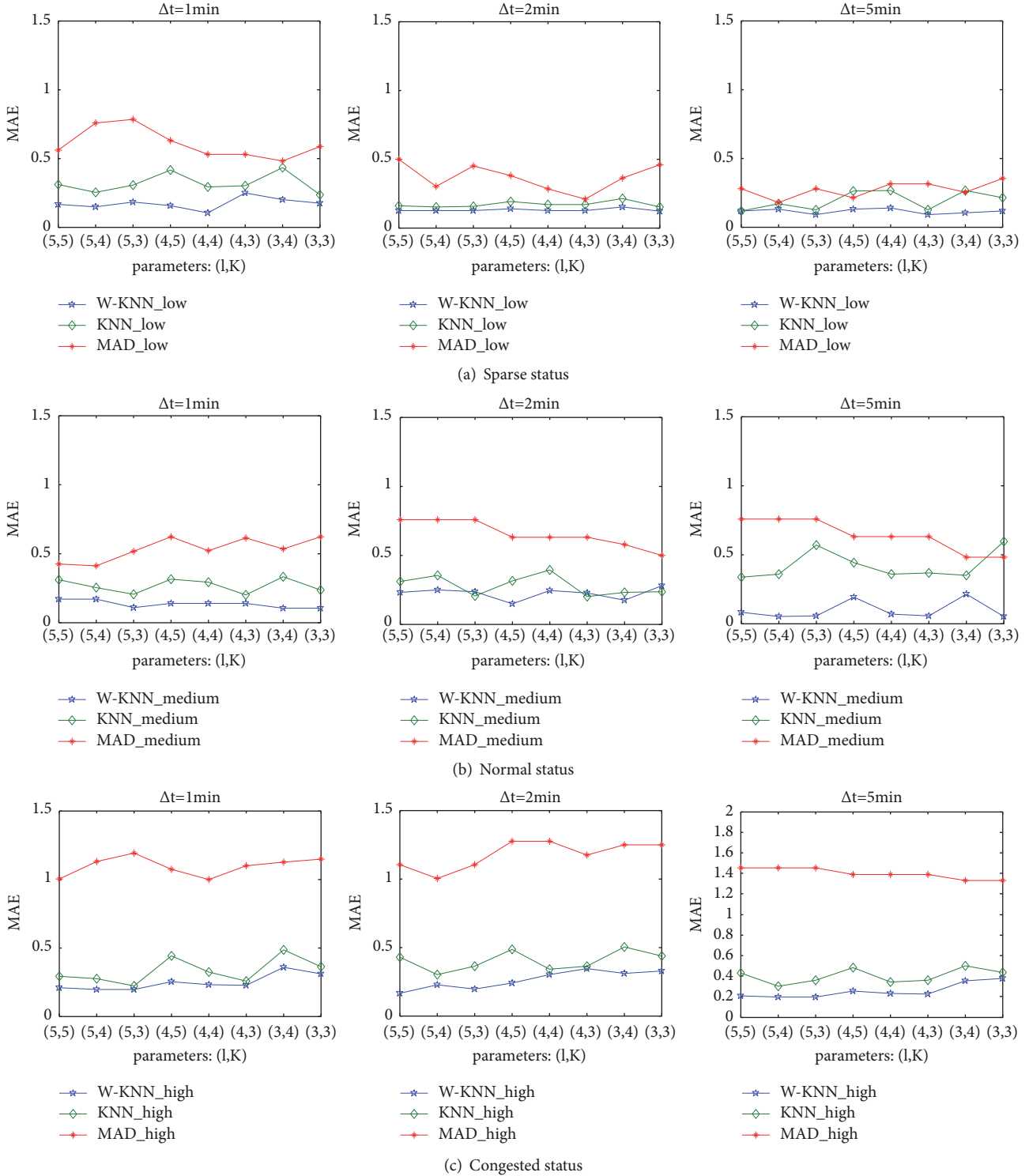


FIGURE 8: Results of three prediction models.

to process real-time traffic information and provide routing decisions dynamically.

In order to connect to a certain vehicle, GPS is leveraged in most routing protocols to locate that node. However, as described in the above, excessive dependence on GPS is

unreliable. Due to the unavoidable defects caused by GPS, researchers are exploring new localization methods avoiding or decreasing the usage of GPS. In [39], authors designed a novel grid-based on-road localization system (GOT), where vehicles with and without accurate GPS signals self-organize

into a VANET. Vehicles in this small size VANET exchange location and distance information and help each other to calculate an accurate position for all the vehicles inside the network. This paper develops the fuzzy geometric relationship among vehicles, and utilizes a novel grid-based mechanism to evaluate the geometric relationships and calculate vehicle locations. Although a light-load grid-based calculation mechanism, which incurs only linear error propagation, is proposed in this paper, there is still a part of vehicles acquiring location information relying on GPS. A GPS-free localization framework that uses two-way time of arrival with partial use of dead reckoning to locate the vehicles based on communication with a single RSU is proposed in [20]. This proposed localization framework consists of two phases, determining the driving direction and computing the vehicle location in the Y-dimension. Compared to existing localization schemes which use multiple RSUs for vehicle localization, this paper decreased the required number of RSUs getting a higher accuracy compared to existing single RSU techniques in the same time. The weakness of this framework is that RSU must be installed at entry or exit. To eliminate those potential troubles caused by GPS and ensuring system scalability, our KNN-based machine learning system serves as a GPS-free dynamic vehicle location prediction system.

Machine learning techniques can learn from training data set automatically to identify rules, and we can use these rules to predict results for testing data. Among all the applications of machine learning methods, classification is one of the most major branches. Classification algorithms are assigned the responsibility of learning an objective function f (20) that maps each set of attributes $\{A_1, A_2, \dots, A_m\}$ to one of the predefined categories or classes $\{C_1, C_2, \dots, C_n\}$ [40].

$$\{A_1, A_2, \dots, A_m\} \xrightarrow{f} C_i \quad 1 \leq i \leq n \quad (20)$$

K-Nearest Neighbor, a simple yet effective classification algorithm, which has been adopted in numerous regression and classification problems, is applied in our location method. As an instance based learning method, KNN classifies each testing data based on a certain amount of instances, and it is based on the principle that the instance within a data set will generally exist in close proximity to other instances that have similar properties. For each testing data, KNN identifies the K-nearest training data from the training data set and stores them in the N_k set. The class of the testing data is same with the majority vote of the N_k set. The only parameter in this algorithm is the number of K neighbors, which can be customized according to concrete applications [41].

In general, machine learning system needs a training process in advance such that it can generalize new instance better. During the training process, interested information will be collected to make up the training data set, and then the training set will serve as input to train the machine learning system after necessary data pre-processing. KNN's high efficiency benefits from its lazy learning characteristic; i.e., we do not need to fix any generic model in advance, so the training phase will be shorter compared with other

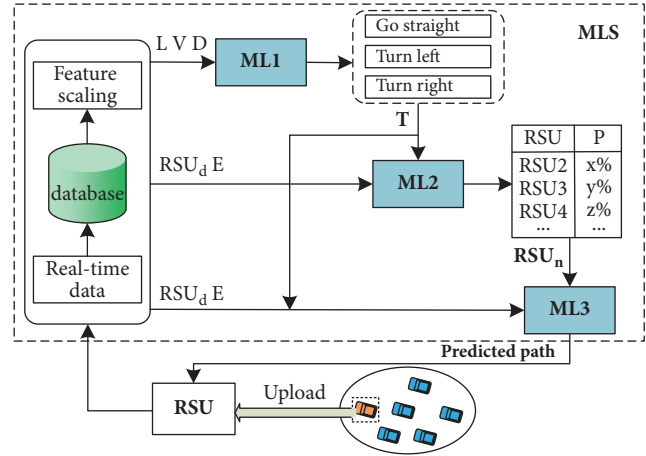


FIGURE 9: Machine learning system.

machine learning algorithms. During the training phase, data collection is done by RSUs and vehicles travelling between these RSUs in coordination with one another. If a vehicle that just left the coverage of RSU_a is entering the coverage of RSU_b , it will transmit its traffic information to RSU_b . According to the received information, RSU_b can inform the previous RSU that vehicle has just travelled through and transmit that information back to RSU_a . The traffic data collected by RSUs compose our training data set and will be retrieved by our machine learning system in the testing phase. Driving features monitored to train the machine learning system include the lane number (L), the vehicle velocity (V), the driving direction (D), the exit of RSU_a (E), the turning direction at the first intersection after leaving RSU_a (T), and the travelling path from the exit E of RSU_a to RSU_b .

After the training process is completed, when a new sample data arrives, machine learning system will make predictions and meanwhile store it as a training data to update the training data set. The designed system can process dynamic traffic information so as to locate the destination vehicle roughly without GPS. Although the system cannot provide accurate position of a vehicle, it can determine which road the predicted vehicle is travelling on precisely, which can lend enough support to making routing decisions for data delivery and eliminate those potential troubles caused by GPS at the same time. For description convenience, in the predicting phase we illustrate an example where the target vehicle has just left RSU_1 from exit 3 and is travelling in the blind zone now. As mentioned, when the destination vehicle was about to leave RSU_1 , it has uploaded its real-time traffic information to RSU_1 . After receiving TREQ from another RSU, RSU_1 will send the latest data of the destination node to its embedded machine learning system database to make predictions. As shown in Figure 9, our machine learning system is composed of three parts.

4.3.1. Predicting Destination's Turning Direction. The first problem that needs to be solved is predicting vehicle's turning direction at intersection, and this will determine the general orientation for our location. In machine learning 1, we select

KNN algorithm to predict which direction the destination vehicle will turn to at next intersection after leaving RSU_d through one of the six exits. In the dynamic prediction process, the target vehicle will first upload its real-time traffic information to RSU_1 . After feature selection, target vehicle's turning direction related training data stored in the database and real-time data just uploaded will be sent into machine learning 1 in RSU_1 together. When approaching an intersection, many elements will influence a vehicle's turning direction more or less. Among all the influencing factors, we select three variables that play leading roles, the lane number (L), the vehicle velocity (V), and the driving direction (D), to predict which category the output should be assigned to. The output data will be one of the three classes: go straight, turn left, and turn right. The nearest neighbors K are set to be 10 which bring the best forecast accuracy according to our simulation results.

4.3.2. Predicting the Probability into Each RSU. Next, our system will determine every potential RSU that destination vehicle might travel to and predict the probability of each potential RSU based on the training data set. In machine learning 2, the predicting outcome from machine learning 1 will be merged with other necessary data uploaded to RSU_1 to predict which RSU coverage area the destination vehicle will move into. The input data will be the previous RSU (RSU_d), the exit of RSU_1 (E), and the predicted turning direction (T). The output will be a table consisting all the possible RSUs as well as their relevant possibilities. Our proposed mechanism will first select the one with the highest possibility to relay packets. If the destination node does not receive packets after a threshold time, RSU with the second highest possibility will be selected and so on. Suppose the predicted turning direction from machine learning 1 is going straight; the most likely RSU that destination vehicle will visit next is RSU_2 as we can learn from Figure 1. Not all vehicles travelling on roads observe the traffic regulations especially in an emergency, so there may be special little results which seem counterintuitive. Vehicles turning around halfway can explain the existence of other RSUs in the table.

4.3.3. Predicting the Travelling Path. Finally, the machine learning system in RSU_1 will predict the travelling path of the destination by locating which road the target vehicle is travelling on. On the basis of the two results obtained above, machine learning 3 will perform this step. As mentioned above, vehicles' running traces are collected into training data set in the training process; therefore, when supplied related information, the database will provide matching paths. Machine learning 3 will take input the exit of RSU_d ($E = 3$), the previous RSU ($RSU_d = RSU_1$), and the predicted turning direction ($T = go\ straight$) and the predicted next RSU ($RSU_n = RSU_2$) to generate output traces from the exit 3 of RSU_1 to RSU_2 . According to the predicted trace, our system can analyze the destination vehicle's location roughly without GPS. After outputting the final outcome to RSU_d , our KNN-based machine learning system has completed a systematic prediction task. Following, the two-way mode transfer will be employed to deliver packets to the destination. Our machine

TABLE 5: Simulation parameters.

Parameters	Value
Simulated area	3 km * 2.5 km
Number of vehicles	[150, 300, 450, 600, 750]
Maximum speed of vehicles	80km/h
RSU communication radius	500 m
Vehicle communication radius	500 m
No. lanes of each direction	2
RSU coverage ratio	41%
Packet size	512 bytes
Vehicle beacon interval	1s

learning system can work as a GPS-free dynamic vehicle location prediction method to acquire the location of vehicle so as to eliminate those potential troubles caused by GPS.

5. Simulation and Evaluation

In this section, we present and discuss the performance of the proposed system through network simulations. In order to evaluate the proposed scheme, we compare it with STAR (Shortest-Path-Based Traffic-Light-Aware Routing) [27] and modified STAR. During the data delivery process, STAR adopts the most common V2V technology; therefore, in terms of wireless transmission, this scheme is representative. When reaching an intersection, STAR attempts to forward packets to a connected red light road segment instead of forwarding packets to the green light road segment. To validate the effectiveness of wired transmission between RSUs, RSUs are involved to deliver packets in the modified STAR.

5.1. Simulation Environment. To simulate the mobility of vehicles and vehicular network, Simulation of Urban Mobility (SUMO) is used for simulating vehicles' mobility traces and road topology, and Network Simulator (NS, version 3.0) is used for simulating vehicular networks. There are six RSUs in our layout as shown in Figure 1, and each RSU is equipped with a dedicated machine learning system. In each simulation experiment, we determine 10 source nodes and 10 destination nodes randomly. Each scheme is tested in the different vehicle densities to analyze their performances when running in various road conditions. Each result in all scenarios is the average of 10 runs. The detailed simulation parameters are shown in Table 5.

5.2. Results and Analysis. To evaluate the performance of these data delivery strategies, three performance metrics are employed: (1) packet delivery ratio, the ratio of the number of the packets successfully received by destination nodes to the total number of the packets sent by source vehicles, (2) network delay, the average latency of the data packets that travel from their source vehicles to the destination vehicles, and (3) control overhead, the number of extra packets generated in the delivery process per minute. In addition, a

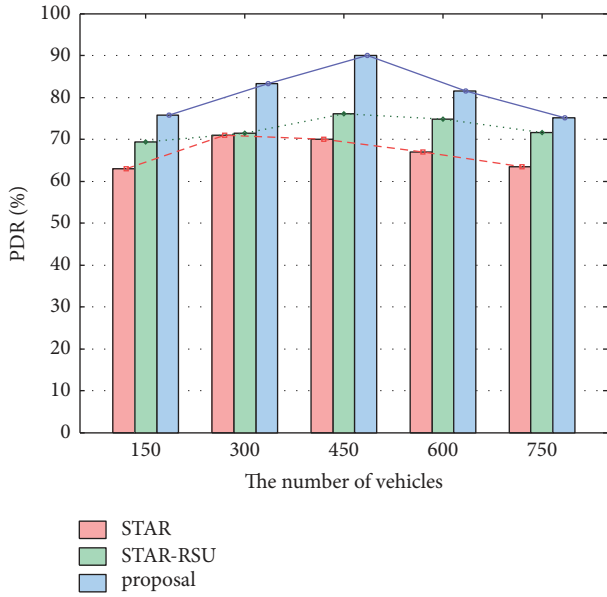


FIGURE 10: Comparison of packet delivery ratio.

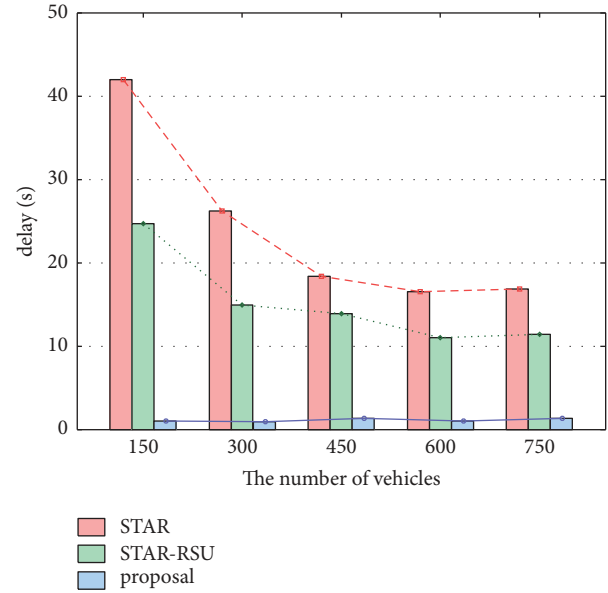


FIGURE 11: Comparison of network delay.

set of simulations were conducted to evaluate the impact of vehicle density.

The variation of packet delivery ratio with different vehicle densities is illustrated in Figure 10. These results illustrate that, with the increase of vehicle density, the packet delivery ratios rise firstly and then decrease in all the three methods. This is because relay vehicles may not be available to establish communication links when the vehicle density is too low, and channel collisions or traffic jams may occur when the vehicle density is too high. The pure STAR suffers the lowest delivery ratio since V2V is the main communication mode. The modified STAR reduces the packet loss delivery by replacing a part of wireless transmission with RSUs, which proves the advantage of utilizing backbone network for vehicular network. Our scheme outperforms the other two strategies throughout the whole running process because our fuzzy-rule-based wireless transmission method optimize V2V communication and the specially designed machine learning system can process dynamic traffic information effectively in different vehicle densities. In our scheme, the packet delivery ratio can be as high as 90% when traffic condition is good (450 vehicles) and can still reach more than 75% even in the worst case (150 vehicles and 750 vehicles).

Figure 11 shows the results of packet delivery delay with three compared schemes in different vehicle densities. The figure shows that, by advancing the number of vehicles from 150 to 750, the average delay for STAR and modified STAR decreases sharply firstly and then levels off. This is because the lower the vehicle density is, the more possible carry-and-forward is adopted, which leads to higher delay. As we compare horizontally, the network delay of our proposal outperforms the other two methods observably and holds steady without obvious fluctuations. The delay in our scheme remains fluctuating around 1 second in the whole process instead of soaring to dozens of seconds as in other two

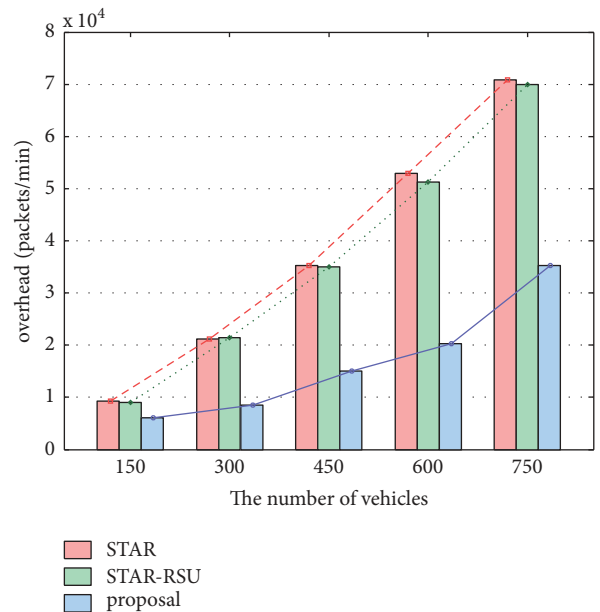


FIGURE 12: Comparison of control overheads.

schemes. This is because the optimized wireless communications and the backbone networks support the fast and efficient transmission of packets. With the participation of RSU, the result of modified STAR is much better than the pure STAR. When running in different vehicle densities, the average delay of STAR and modified STAR varied intensely, instead of keeping stable in our scheme, which means that our proposed method can adapt to different traffic conditions.

Figure 12 compares the control overhead among different schemes to evaluate the costs. As the figure shows, with the increase of the vehicle density, the overheads rise for all three strategies. However, the results of the proposed

strategy are much better than the other two methods in the whole simulation experiment. When the number of vehicles is greater than or equal to 300, overheads in STAR and modified STAR are twice or more that in our scenario. In STAR, both delivering packets in V2V communication mode and sending messages between intersections to check the connectivity will increase the cost significantly. Compared with STAR, the modified STAR decreases overhead slightly with the assistance of RSU in the delivery process. In contrast, our proposal causes the least overhead, because the teamwork of machine learning system and RSU will analyze traffic data timely to determine the delivery path instead of sending a great deal of communication signals between vehicles. And the vehicle-based short-term speed prediction method makes the relay node selection more reliable so as to reduce overhead caused by building communication links continually.

In summary, our proposed data delivery scheme can improve the packet delivery ratio, guarantee the timeliness of messages, and reduce the control overhead significantly. Consequently, our proposed scheme is suitable for data delivery in urban scenario.

6. Conclusion

In this paper, we propose a novel data delivery scheme for vehicular networks in urban environments, and we focus on the analysis that both the source node and the destination node are in the blind zone.

In order to set up delivery paths for vehicles in the blind zone, we designed a fuzzy-rule-based wireless transmission method. This key technology will select an optimal option from all the possible paths with considering multiple factors comprehensively. By optimizing V2V communications, the DTQ can be improved. One of the key technologies in this fuzzy-rule-based approach is the vehicle speed prediction approach. Different from common segment-based prediction method, we designed a vehicle-based short-term vehicle speed prediction method. Taking full consideration of velocity by comparing predicted speeds at next moment will make the relay node selection more reliable and increase the stability of selected transmission link. Another key technology in our data delivery scheme is the specially designed machine learning system embedded in each RSU, which provides routing decisions by processing dynamic traffic information delivered to it. The combination of machine learning system and RSU empowers our system to abandon GPS without degrading the network performance. The wired communication between RSUs can reduce the delay resulted from the unreliable carry-and-forward manner in the pure V2V communication network.

The performance of our proposal has been verified through simulations in NS-3. For future work, we intend to apply more machine learning methods to the study of VANETs.

Conflicts of Interest

The authors declare that they have no conflicts of interest.

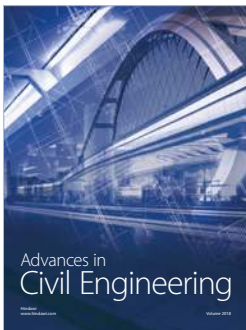
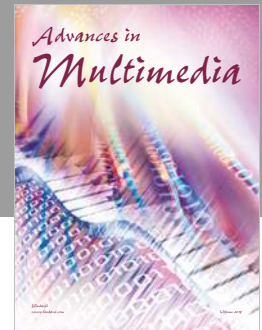
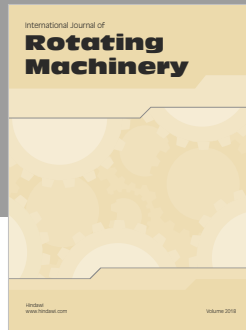
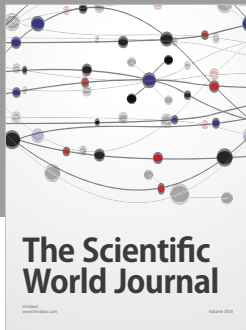
Acknowledgments

This work was supported by Henan International Science & Technology Cooperation Program (182102410050), Henan Young Scholar Promotion Program (2016GGJS-018), the Program for Science & Technology Development of Henan Province (162102210022), Key Project of Science and Technology Research of the Education Department of Henan Province (17A413001), and CERNET Innovation Project (NGII20151005).

References

- [1] W. Quan, Y. Liu, H. Zhang, and S. Yu, "Enhancing crowd collaborations for software defined vehicular networks," *IEEE Communications Magazine*, vol. 55, no. 8, pp. 80–86, 2017.
- [2] N. Cheng, N. Zhang, N. Lu, X. Shen, J. W. Mark, and F. Liu, "Opportunistic spectrum access for CR-VANETs: A game-theoretic approach," *IEEE Transactions on Vehicular Technology*, vol. 63, no. 1, pp. 237–251, 2014.
- [3] N. Cheng, F. Lyu, and J. Chen, *Big Data Driven Vehicular Networks*, IEEE Network, 2018.
- [4] C. Silva, M. Nogueira, D. Kim, E. Cerqueira, and A. Santos, "Cognitive radio based connectivity management for resilient end-to-end communications in VANETs," *Computer Communications*, vol. 79, pp. 1–8, 2016.
- [5] J.-S. Li, I.-H. Liu, C.-K. Kao, and C.-M. Tseng, "Intelligent Adjustment Forwarding: A compromise between end-to-end and hop-by-hop transmissions in VANET environments," *Journal of Systems Architecture*, vol. 59, no. 10, pp. 1319–1333, 2013.
- [6] I. Tal and G. Muntean, "Towards Reasoning Vehicles," *ACM Computing Surveys*, vol. 50, no. 6, pp. 1–37, 2018.
- [7] N. Cheng, N. Lu, N. Zhang, X. S. Shen, and J. W. Mark, "Vehicular WiFi offloading: Challenges and solutions," *Vehicular Communications*, vol. 1, no. 1, pp. 13–21, 2014.
- [8] X. Xiang, X. Wang, and Z. Zhou, "Self-Adaptive On-Demand Geographic Routing for Mobile Ad Hoc Networks," *IEEE Transactions on Mobile Computing*, vol. 11, no. 9, pp. 1572–1586, 2012.
- [9] W. K. Lai, M. Lin, and Y. Yang, "A Machine Learning System for Routing Decision-Making in Urban Vehicular Ad Hoc Networks," *International Journal of Distributed Sensor Networks*, vol. 11, no. 3, article 374391, pp. 1–13, 2015.
- [10] S. Agrawal, R. S. Raw, N. Tyagi, and A. K. Misra, "Fuzzy logic based greedy routing (flgr) in multi-hop vehicular adhoc networks," *Indian Journal of Science and Technology*, vol. 8, no. 30, pp. 1–14, 2015.
- [11] A. Jahangiri and H. A. Rakha, "Applying Machine Learning Techniques to Transportation Mode Recognition Using Mobile Phone Sensor Data," *IEEE Transactions on Intelligent Transportation Systems*, vol. 16, no. 5, pp. 2406–2417, 2015.
- [12] H. Zhou, N. Cheng, Q. Yu, X. S. Shen, D. Shan, and F. Bai, "Toward multi-radio vehicular data piping for dynamic DSRC/TVWS spectrum sharing," *IEEE Journal on Selected Areas in Communications*, vol. 34, no. 10, pp. 2575–2588, 2016.
- [13] Z. Su, Y. Hui, and Q. Yang, "The next generation vehicular networks: a content-centric framework," *IEEE Wireless Communications Magazine*, vol. 24, no. 1, pp. 60–66, 2017.
- [14] O. K. Tonguz, N. Wisitpongphan, J. S. Parikh, F. Bai, P. Mudalige, and V. K. Sadekar, "On the Broadcast Storm Problem in Ad hoc Wireless Networks," in *Proceedings of the 2006 3rd International*

- Conference on Broadband Communications, Networks and Systems*, pp. 1–11, San Jose, CA, USA, October 2006.
- [15] D. Ou, Y. Yang, L. Xue, T. Shen, and L. Zhang, “The analytic hierarchy process-based optimal forwarder selection in multi-hop broadcasting scheme for vehicular safety,” in *Proceedings of the 2014 IEEE Intelligent Vehicles Symposium (IV)*, pp. 999–1004, MI, USA, June 2014.
- [16] A. T. Reza, T. A. Kumar, and T. Sivakumar, “Position Prediction based Multicast Routing (PPMR) using Kalman filter over VANET,” in *Proceedings of the 2nd IEEE International Conference on Engineering and Technology, ICETECH 2016*, pp. 198–206, India, March 2016.
- [17] A. M. Vegni and E. Natalizio, “Forwarder smart selection protocol for limitation of broadcast storm problem,” *Journal of Network and Computer Applications*, vol. 47, pp. 61–71, 2015.
- [18] L. A. Zadeh, “Fuzzy sets,” *Information and Computation*, vol. 8, pp. 338–353, 1965.
- [19] P. Lopez-Garcia, E. Onieva, E. Osaba, A. D. Masegosa, and A. Perallos, “A hybrid method for short-term traffic congestion forecasting using genetic algorithms and cross entropy,” *IEEE Transactions on Intelligent Transportation Systems*, vol. 17, no. 2, pp. 557–569, 2016.
- [20] A. A. Wahab, A. Khattab, and Y. A. Fahmy, “Two-way TOA with limited dead reckoning for GPS-free vehicle localization using single RSU,” in *Proceedings of the 13th International Conference on ITS Telecommunications (ITST '13)*, vol. 12(1), pp. 244–249, Tampere, Finland, November 2013.
- [21] B. Cheng, G. Liao, C. Chen, and H. Chen, “A Multi-level Fuzzy Comprehensive Evaluation Approach for Message Verification in VANETs,” in *Proceedings of the 2012 Third FTRA International Conference on Mobile, Ubiquitous, and Intelligent Computing (MUSIC)*, pp. 176–181, Vancouver, Canada, June 2012.
- [22] A. Afful-Dadzie, E. Afful-Dadzie, S. Nabareseh, and Z. K. Oplatková, “Tracking progress of African Peer Review Mechanism (APRM) using fuzzy comprehensive evaluation method,” *Kybernetes*, vol. 43, no. 8, pp. 1193–1208, 2014.
- [23] T. L. Saaty, *The Analytic Hierarchy Process*, McGraw-Hill, New York, NY, USA, 1980.
- [24] K. Katsaros, M. Dianati, Z. Sun, and R. Tafazolli, “An evaluation of routing in vehicular networks using analytic hierarchy process,” *Wireless Communications and Mobile Computing*, vol. 16, no. 8, pp. 895–911, 2016.
- [25] M. Patra, S. Mishra, and C. S. Murthy, “An Analytic Hierarchy Process Based Approach for Optimal Road Side Unit Placement in Vehicular Ad Hoc Networks,” in *Proceedings of the 2014 IEEE Vehicular Technology Conference (VTC 2014-Spring)*, pp. 1–5, Seoul, South Korea, May 2014.
- [26] N. V. Dharani Kumari and B. S. Shylaja, “AMGRP: AHP-based Multimetric Geographical Routing Protocol for Urban environment of VANETs,” *Journal of King Saud University - Computer and Information Sciences*, 2017.
- [27] J.-J. Chang, Y.-H. Li, W. Liao, and I.-C. Chang, “Intersection-based routing for urban vehicular communications with traffic-light considerations,” *IEEE Wireless Communications Magazine*, vol. 19, no. 1, pp. 82–88, 2012.
- [28] B. Yao, C. Chen, Q. Cao et al., “Short-term traffic speed prediction for an urban corridor,” *Computer-Aided Civil and Infrastructure Engineering*, vol. 32, no. 2, pp. 154–169, 2016.
- [29] H. Chang, Y. Lee, B. Yoon, and S. Baek, “Dynamic near-term traffic flow prediction: System-oriented approach based on past experiences,” *IET Intelligent Transport Systems*, vol. 6, no. 3, pp. 292–305, 2012.
- [30] S. Clark, “Traffic prediction using multivariate nonparametric regression,” *Journal of Transportation Engineering*, vol. 129, no. 2, pp. 161–168, 2003.
- [31] B. L. Smith and M. J. Demetsky, “Multiple-interval freeway traffic flow forecasting,” *Transportation Research Record*, no. 1554, pp. 136–141, 1996.
- [32] H. Chang, D. Park, S. Lee, H. Lee, and S. Baek, “Dynamic multi-interval bus travel time prediction using bus transit data,” *Transportmetrica*, vol. 6, no. 1, pp. 19–38, 2010.
- [33] P. Kang and S. Cho, “Locally linear reconstruction for instance-based learning,” *Pattern Recognition*, vol. 41, no. 11, pp. 3507–3518, 2008.
- [34] D. P. Solomatine, M. Maskey, and D. L. Shrestha, “Instance-based learning compared to other data-driven methods in hydrological forecasting,” *Hydrological Processes*, vol. 22, no. 2, pp. 275–287, 2008.
- [35] C. G. Atkeson, A. W. Moore, and S. Schaal, “Locally weighted learning,” *Artificial Intelligence Review*, vol. 11, no. 1–5, pp. 11–73, 1997.
- [36] D. W. Aha and R. L. Goldstone, “Concept learning and flexible weighting,” in *Proceedings of 14th Annual Conference of the Cognitive Science Society*, pp. 534–539, Mahwah, USA, 1992.
- [37] M. P. Wand and W. R. Schucany, “Gaussian-based kernels,” *Canadian Journal of Statistics*, vol. 18, no. 3, pp. 197–204, 1990.
- [38] V. V. Fedorov, P. Hackl, and W. G. Müller, “Moving local regression: the weight function,” *Journal of Nonparametric Statistics*, vol. 2, no. 4, pp. 355–368, 1993.
- [39] T. Yan, W. Zhang, and G. Wang, “A grid-based on-road localization system in VANET with linear error propagation,” *IEEE Transactions on Wireless Communications*, vol. 13, no. 2, pp. 861–875, 2014.
- [40] L. P. Portugal-Poma, C. A. Marcondes, H. Senger, and L. Arantes, “Applying Machine Learning to Reduce Overhead in DTN Vehicular Networks,” in *Proceedings of the 2014 Brazilian Symposium on Computer Networks and Distributed Systems (SBRC)*, pp. 94–102, Florianopolis, Brazil, May 2014.
- [41] S. B. Kotsiantis, “Supervised machine learning: a review of classification techniques,” in *Proceedings of the Conference on Emerging Artificial Intelligence Applications in Computer Engineering: Real World Ai Systems with Applications in Ehealth, Hci, Information Retrieval and Pervasive Technologies IOS Press*, pp. 3–24, 2007.



Hindawi

Submit your manuscripts at
www.hindawi.com

