International Journal of Future Generation Communication and Networking
Vol. 7, No. 6 (2014), pp. 217-230
http://dx.doi.org/10.14257/ijfgcn.2014.7.6.20

Logical Connectivity Prediction Models for VANET based on Nonlinear Regression and ELM: An Example of the AODV Protocol

Haiqing Liu, Licai Yang, Sijing Ding and Yunfeng Shi

School of control science and engineering, Shandong University, Jinan, 250061, Shandong, P.R. China
hqliu623@gmail.com, yanglc@sdu.edu.cn, dsj9005@gmail.com and yantaiyunfeng@126.com

Abstract

Differing from the physical connectivity of the topology structure, the logical connectivity of VANET considers both the interior network configuration and the external communication environment. Hence, the traditional mathematical analysis and modeling methods which are usually used in physical connectivity research are no longer suitable for the logical connectivity prediction. Taking the AODV protocol as an example, this paper simulates the effects of different road traffic parameters on logical connectivity probability and selects three main effect factors, roadway length, vehicle number and vehicle speed. Furthermore, the inner relation between the logical connectivity and the three road traffic parameters is studied based on data mining technique and then two logical connectivity prediction models are presented, the nonlinear regression-based model and the extreme learning machine-based model. Simulation results show that the two models are both with high accuracy in predicting the network logical connectivity under different road traffic environments.

Keywords: *VANET, logical connectivity, nonlinear regression, ELM, AODV*

1. Introduction

As a special type of mobile ad hoc networks (MANETs), vehicular ad hoc networks (VANETs) allow vehicles to form a self-organized network without the requirement of permanent infrastructures. It is a promising application-oriented network deployed along a roadway for 1) vehicular safety-related applications such as collision warning systems, road condition warning, lane-changing assistance [1, 2], 2) transportation efficiency-related applications such as traffic light control, vehicle navigation [3, 4] and 3) entertainment applications. Unlike conventional ad hoc wireless types of networks, a VANET may be required to deal with some new issues caused by the road traffic environment. For example, the vehicles move in high speeds randomly, causing dynamic and rapidly changing topologies of VANET. Further, because of obstacles to wireless signal by large objects, e.g. skyscrapers in cities, communications between vehicles must have line-of-sight and the paths are confined in the zonal roadway structure. What's more, influenced by human driving behavior and the traffic regulations, the distribution of vehicles on roads is uneven, and it is hard to be modeled using a simple mathematical method.

The peculiar characteristics of VANET bring great challenges to a real-time, efficient, reliable transmission of messages. The multi-hop communication network connectivity is a fundamental performance measure when transmitting messages in VANET. Two vehicles in the network are connected if they can exchange information with each other, either directly or indirectly [5]. Employing the ideas of network layering, VANET connectivity research can be

ISSN: 2233-7857 IJFGCN

classified into two cases: physical connectivity and logical connectivity [6].The physical connectivity refers to the topology connection between two nodes at the physical layer. In practical application, it can be tested by sending and receiving electrical level. Good physical connectivity is the prerequisite for providing reliable service to network users and the basis for the logical connectivity. While the logical connectivity implies the reachability of messages when running MAC protocol, routing protocol and topology control mechanism in the network under the condition that the physical topology is connected. Generally, the logical connectivity can be described by the packet loss probability index.

VANET is a typical self-organizing network, the performance of which is highly dependent on the road traffic conditions. The point-to-point data transmission is affected not only by the configurations, protocols and bandwidth of the network, but also by the characteristics of roadway and the traffic flow, such as roadway length, lane number, vehicle number, vehicle velocity and vehicle density. In this paper, taking the AODV protocol as an example, we simulate and analyze the relationship between the VANET logical connectivity probability and the traffic parameters. Further, using two different data mining technologies, the nonlinear regression and the extreme learning machine (ELM), we present two logical connectivity prediction models according to three main effect factors which are roadway length, vehicle number and speed. Simulation results show that, given certain traffic parameters, the two proposed models can forecast the logical connectivity of VANET accurately.

The rest of the paper is organized as follows. Section 2 exhibits the related work on connectivity research in VANET. In Section 3, the simulation data acquisition process is presented and the main effect factors on the connectivity are analyzed. In Section 4, the logical connectivity prediction models based on nonlinear regression and ELM are presented respectively. The performed simulations together with the result evaluations are presented in Section 5. In Section 6, we conclude the paper.

2. Related Work

Network connectivity is one of the most important issues in VANETs to ensure reliable dissemination of time-critical information. While in the road traffic communication environment, the dynamic and rapidly changing topologies of vehicular networks can cause frequent link disconnections, which has been a bottleneck in designing protocols and transmitting messages in VANET [7]. In recent years, many scholars and institutions have devoted to the VANET connectivity research and a number of studies concerning VANET connectivity modeling and analysis have been reported.

The existing research of VANET connectivity can be classified into two aspects: connectivity analysis and connectivity prediction. The connectivity analysis research aims to search for the main effect factors on the logical connectivity and how these factors affect it, and the necessary conditions of a completely-connected network. The connectivity prediction research is to establish connectivity prediction models for the network performance evaluation and the routing protocol design under different road traffic environments, vehicle movements and distributions, and network configurations.

Paper [6] provides a probability analysis algorithm to calculate the necessary condition of 1-connected VANET in highway scenarios. It makes a conclusion that the radio communication range of each node must be subject to $\Theta(|\log(1-p^{1/n})|/n)$ in order to ensure that there is no isolated node within the entire network. Further, in [8], authors investigate the k-connectivity of 1-D linear vehicular ad hoc network based on matrix of decomposition. The analysis indicates that the expectation of the maximum number of tolerable vehicle departures almost linearly increases with the total number of vehicles.

In [9], authors investigate the VANET connectivity properties from a physical layer perspective. The minimum transmitting power used by all vehicles, sufficient to guarantee network connectivity, is studied. As opposed to the conventional graph-theoretic approach, the network connectivity problem is analyzed according to a physical layer-based quality of service constraint. The analysis provides a frame work for investigating the impact of traffic dependent parameters such as vehicle arrival rate, vehicle density, mean and standard deviation of vehicle speed, highway length and physical layer based parameters.

In [10], authors present an analytical model for multi-hop connectivity of IVC in a traffic stream, in which positions of vehicles are no longer described by certain geometric distributions, but through observations, traffic simulators or traffic theories. They derive a recursive model of node and hop probabilities and define a number of performance measures of multi-hop connectivity. The model is applied to study multi-hop connectivity of IVC in both uniform and non-uniform traffic. The proposed model is efficient without repeating traffic simulations while capable of capturing the impact of arbitrary distribution patterns of vehicles and is suitable for evaluating connectivity of IVC for different traffic congestion patterns.

In [11], authors propose a cell-based connectivity model for VANET in a road segment. The road segment, which is a portion of a street between two adjacent intersections, is divided into several cells, the length of which is set as the average length of vehicles. Each cell can obtain at most one vehicle and each vehicle can occupy only one cell. With the assumption of uniformly distributed vehicles, the probability that the network is disconnected equals to the probability that the length of successive empty cells on the road segment is longer than the communication range. Given the parameters of roadway length, number of lanes and vehicles, network connectivity probability of the road segment can be calculated using the model.

Summarizing the literatures mentioned above, we conclude the current status and the existing issues of VANET connectivity study from the major research contents and the methods. On one hand, the main research work concentrates on the physical connectivity analysis and modeling. Actually, influenced by the node mobility, traffic environment and the routing protocol performance, messages cannot be guaranteed to reach the target node even though the physical topology is connected. While the logical connectivity research takes into account all the possible effect factors when transmitting messages and directly determines the messages reachable or not. Hence the logical connectivity study is also necessary in VANET. On the other hand, the mainstream research methods of VANET connectivity include graph theory based, probability analysis based, statistics based [12] and percolation theory based [13] methods. These methods are usually applied on the basis of some assumptions of vehicle distributions and movements. For example, vehicles on the road are distributed following some geometric distribution, the Poisson distribution or uniform distribution. The movement of vehicles mainly adopts the random way point (RWP) model. However, influenced by the human driving behavior and the traffic regulations, these assumptions cannot reflect the distribution and movement of vehicles on road truly. Therefore, the current VANET connectivity research methods have more or less limitations.

3. Simulation Data Acquisition and Effect Factors Analysis

Before introducing our work, we first make some necessary and reasonable assumptions:

(1) Roadways which we study are bi-directional and have the same lane number for the two directions.

(2) All vehicles are equipped with wireless communication devices and they can send messages to each other.

(3) Vehicles on the roadway are in the status of free flow, regardless of the impacts caused by traffic lights.

(4) Wireless transmission range is definite. The reason is that once VANET and the corresponding applications are commercialized, the physical criterion of wireless communication devices would be standardized. Wireless transmission range can be considered as a definite value.

3.1. Simulation Data Acquisition

Besides the internal network configuration, the logical connectivity of VANET is mainly concerned with the external communication environment, which can be shown in two aspects: the roadway physical properties and distribution of vehicles. The roadway physical properties include length and lane number. The distribution of vehicles can be described by the macro statistical characteristics of traffic flow, such as vehicle number, vehicle speed and density. In this paper, the logical connectivity research aims to seek the internal relationship between the connectivity probability and road traffic factors under certain protocol condition.

In order to acquire actual data for the connectivity analysis, a large number of simulation experiments are done by the use of VanetMobiSim/NS-2. The vehicle movement model we use in the simulation is the Intelligent Driver Model with Lane Changes (IDM_LC). In the model, vehicles can implement lane change behavior in reality according to the acceleration variation. Using the model, an approximate real road traffic scenario will be simulated and the unreasonable assumptions in the traditional analytical model will be avoided. Parameters of the road traffic and the network configuration are shown in Table 1 and Table 2 respectively.

Table 1. Parameters of Road Way and Vehicles

Parameters	Value
Maximum acceleration(m/s^2)	0.6
Comfortable deceleration(m/s^2)	0.9
Maximum safe deceleration(m/s^2)	4
Lane changing acceleration threshold(m/s^2)	0.2
Roadway length(m)	600, 800, 1000, 1200, 1400, 1600, 1800, 2000
Vehicle number	10, 15, 20, 25, 30,35, 40, 45, 50
Vehicle speed(m/s)	5, 10, 15, 20, 25, 30,35, 40, 45, 50
Lane number(bi-directional)	2, 4, 6

Table 2. Parameters of Network Configuration

Parameters	Values
Propagation model	TwoRayGround
MAC protocol	IEEE 802.11
Routing protocol	AODV
Maximum segment size	50
Communication range (m)	250
Simulation time(s)	2000

Under the conditions with different parameters shown in Table 1, the packet loss rate is simulated. Each condition is simulated 10 times independently. The average packet loss rate of the 10 times simulation p' is calculated as the result of each condition. Let $p = 1 - p'$ as the measure indicator of logical connectivity probability.

3.2. Effects of Vehicle Density on Logical Connectivity Probability

The vehicle density can be calculated by (1), where n is the vehicle number and l is the roadway length.

$$\rho = n / l \tag{1}$$

In the traditional physical connectivity research, the connectivity probability is usually considered as a definite value in case the vehicle density is constant. Generally, this kind of conclusion is acquired based on the assumption that the locations of vehicles follow some certain geometric distributions. However, influenced by driving behavior such as vehicle queuing, lane changing and vehicle following, the vehicle distributions are always different on the roadways which have different lengths and vehicle numbers but the same vehicle density. Figure 1 shows the connectivity probabilities where the vehicle density is 12.5/km but roadway length, vehicle speed and vehicle number are different.

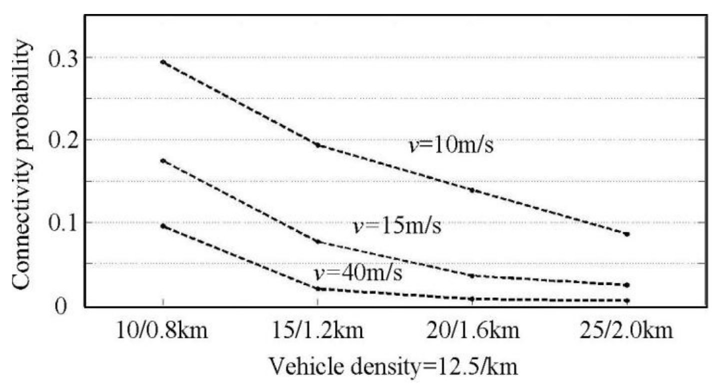


Figure 1. Connectivity Probabilities under the Same Vehicle Density but Different Roadway Lengths, Vehicle Speeds and Vehicle Numbers

From the results, we can see that the connectivity probabilities are different even though the vehicle density is constant. With the increase of roadway length and vehicle number, the connectivity probability decreases. That’s because the larger the roadway length and the vehicle number is, the more probability of driving behavior occurs, and the more frequently the topology changes. Therefore, the vehicle density index which is usually used in physical connectivity research is no longer suitable for logical connectivity analysis. Since the vehicle density can be seen as a correlation value of roadway length and vehicle number, we use these two parameters instead of the density index for logical connectivity analysis in our paper.

3.3. Effects of Lane Number on Logical Connectivity Probability

Since the communication range is far larger than lane width, the connectivity of VANET will be weakly influenced by the lane number. In order to verify this conclusion, this paper simulates the network connectivity under different roadway length, vehicle number and vehicle speed as the bi-directional lane numbers are 2, 4 and 6 respectively.

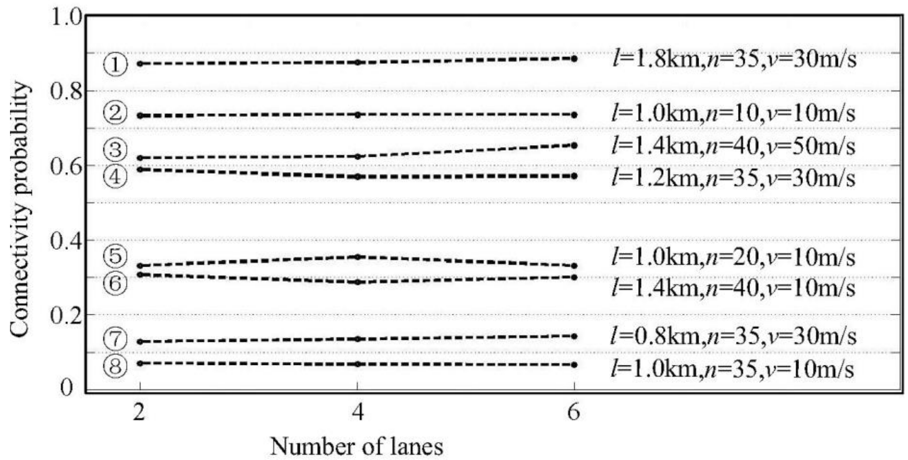


Figure 2. Connectivity Probabilities under Different Lane Numbers

Figure 2 shows the connectivity probability under different lane numbers as 1) the vehicle number and vehicle speed are constant, roadway length changes (line ①, ④, ⑦), 2) the roadway length and vehicle speed are constant, vehicle number changes (line ②,⑤,⑧), 3) the roadway length and vehicle number are constant, vehicle speed changes (line ③,⑥). From the results we can conclude that the connectivity probabilities are almost the same in different lane number cases. The lane number index almost makes no effect to the logical connectivity of VANET. In order to simplify work, we ignore the effects of lane number on connectivity probability in this paper.

3.4. Effects of Roadway Length, Vehicle Number and Vehicle Speed on Logical Connectivity Probability

The effects of roadway length, vehicle speed and vehicle number on connectivity probability are show in Figure 3, where the speed changes from 5 to 55 (m/s), vehicle number changes from 10 to 50 and roadway lengths are 800, 1200, 1600, 2000 (m) respectively.

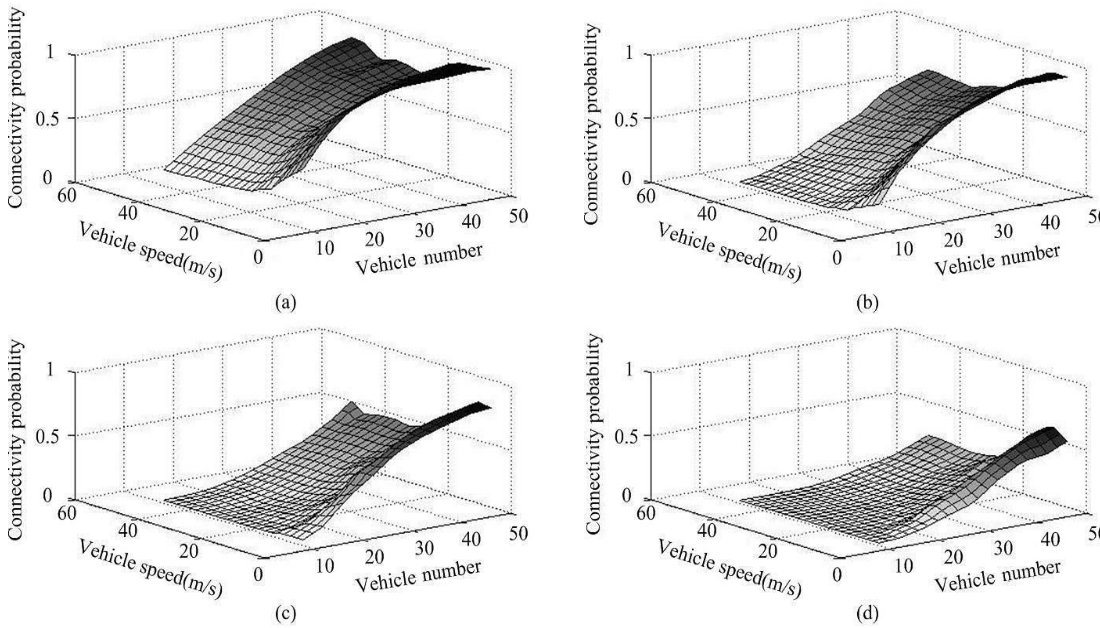


Figure 3. Connectivity Probabilities Distribution under Different Speeds, Vehicle Numbers and Roadway Lengths ((a) Roadway Length=800m, (b) Roadway Length=1200m, (c) Roadway Length=1600m, (d) Roadway Length=2000m)

From the simulation results we can conclude that the connectivity is positive correlation with vehicle number and negative correlation with roadway length and vehicle speed.

As previous analysis, we select three key influencing factors, roadway length, vehicle number and vehicle speed, as the main parameters to explore the statistical regularity of logical connectivity of VANET. Construct the relationship function between the dependent value (connectivity probability (p)) and three independent values (roadway length (l), vehicle number (n) and vehicle speed (v)), as is shown by (2)

$$p = f(l, n, v) \tag{2}$$

Since the logical connectivity takes into account all possible factors both in the internal network configuration and the external communication environment, the traditional analysis methods used in physical connectivity research are no longer suitable. Based on large amount of simulation data, we build connectivity prediction models using two different data mining technologies, the nonlinear regression and the extreme learning machine respectively.

3. Logical Connectivity Prediction Models based on Nonlinear Regression and ELM

3.1. Connectivity Prediction Model based on Nonlinear Regression

There is a certain nonlinear correlation between connectivity and the three parameters. In this section, using the nonlinear regression method, we seek the internal relation and build a simulation model for connectivity prediction.

In statistics, nonlinear regression is a form of regression analysis in which observational data are modeled by a function which is a nonlinear combination of the model parameters and depends on one or more independent variables. Since the effect of the independent variables on connectivity probability is nonlinear, the logical connectivity prediction belongs to the technical field of multivariate nonlinear regression analysis. Making convenience of computation, a cubic regression equation is used to express the mathematical model of the connectivity probability, as (3).

$$p = f(l,n,v) = \sum_{i=0}^{3}\sum_{j=0}^{3-i}\sum_{k=0}^{3-i-j} a_{i,j,k} l^i n^j v^k \text{ , } (i=0,\cdots,3,\ j=0,\cdots,3-i, k=0,\cdots,3-i-j) \quad (3)$$

Where $a_{i,j,k}$ is the coefficient.

The aim for nonlinear regression is to calculate the coefficient matrix in (3) by the sample of the simulation connectivity values. According to the least square principle, the regression equation takes the minimum sum of squared residuals of the predicted values and the simulated values as the fitting criterion, which can be expressed by (4).

$$\min e^2 = \sum_{t=1}^{N}(\sum_{i=0}^{3}\sum_{j=0}^{3-i}\sum_{k=0}^{3-i-j} a_{i,j,k} l_t^i n_t^j v_t^k - \hat{p}_t)^2 \quad (4)$$

Where $\hat{p}_t$ is the simulated value of connectivity probability.

According to the multivariate function extreme value theorem, the necessary condition to gain a minimum sum of squared residuals is:

$$\frac{\partial e^2}{\partial a_{i,j,k}} = 0 \quad (5)$$

Equation (5) can be expressed as (6) and further transformed as (7)

$$2\sum_{t=1}^{N}(\sum_{i=0}^{3}\sum_{j=0}^{3-i}\sum_{k=0}^{3-i-j} a_{i_0,j_0,k_0} l_t^i n_t^j v_t^k - \hat{p}_t) \bullet l_t^{i_0} n_t^{j_0} v_t^{k_0} = 0 \quad (6)$$

$$\sum_{t=1}^{N}\sum_{i=0}^{3}\sum_{j=0}^{3-i}\sum_{k=0}^{3-i-j} a_{i_0,j_0,k_0} l_t^i n_t^j v_t^k \bullet l_t^{i_0} n_t^{j_0} v_t^{k_0} = \sum_{t=1}^{N} \hat{p}_t \bullet l_t^{i_0} n_t^{j_0} v_t^{k_0} \quad (7)$$

Equation (7) can be expressed by a linear matrix equation, as is shown by (8).

$$\begin{bmatrix} \sum_{t=1}^{N}\varphi_t^{0,0,0}\bullet\varphi_t^{0,0,0} & \sum_{t=1}^{N}\varphi_t^{0,0,0}\bullet\varphi_t^{0,0,1} & \cdots & \sum_{t=1}^{N}\varphi_t^{0,0,0}\bullet\varphi_t^{3,0,0} \\ \sum_{t=1}^{N}\varphi_t^{0,0,1}\bullet\varphi_t^{0,0,0} & \sum_{t=1}^{N}\varphi_t^{0,0,1}\bullet\varphi_t^{0,0,1} & \cdots & \sum_{t=1}^{N}\varphi_t^{0,0,1}\bullet\varphi_t^{3,0,0} \\ \vdots & \vdots & \ddots & \vdots \\ \sum_{t=1}^{N}\varphi_t^{3,0,0}\bullet\varphi_t^{0,0,0} & \sum_{t=1}^{N}\varphi_t^{3,0,0}\bullet\varphi_t^{0,0,1} & \cdots & \sum_{t=1}^{N}\varphi_t^{3,0,0}\bullet\varphi_t^{3,0,0} \end{bmatrix} \begin{bmatrix} a_{0,0,0} \\ a_{0,0,1} \\ \vdots \\ a_{3,0,0} \end{bmatrix} = \begin{bmatrix} \sum_{t=1}^{N}\hat{p}_t\bullet\varphi_t^{0,0,0} \\ \sum_{t=1}^{N}\hat{p}_t\bullet\varphi_t^{0,0,1} \\ \vdots \\ \sum_{t=1}^{N}\hat{p}_t\bullet\varphi_t^{3,0,0} \end{bmatrix} \quad (8)$$

Where

$$\varphi_t^{i,j,k} = l_t^i n_t^j v_t^k \quad (9)$$

Let C represent the coefficient matrix in equation (9), $C=[a_{0,0,0}, a_{0,0,1}, \cdots, a_{3,0,0}]^{\mathrm{T}}$. Given N samples of the connectivity probabilities under different roadway length, vehicle number and vehicle speed, C can be calculated and then regression equation (3) can be established.

3.2. Connectivity Prediction Model based on ELM

The artificial neural network is an important method for data mining. The traditional Back-Propagation Network (BP Network) adjusts the weight parameters by the gradient descent iterative algorithm. It has apparent defects: 1) Slow training speed and long computation time, 2) Easy to run into the local least value, 3) Poor generalization ability caused by overtraining. To overcome these defects, Huang [14, 15] proposes the extreme learning machine algorithm based on Moore-Penrose generalized inverse matrix. As a new learning algorithm for single-hidden layer feed-forward neural network, the ELM doesn't need to adjust the input weights and the hidden layer biases. It can obtain the unique optimal solution just by one step calculation. The learning algorithm has advantages as fast learning speed and good generalization performance.

The network training model of ELM employs single-hidden layer feed-forward structure. Let m, M, n be the numbers of nodes of input layer, hidden layer and output layer respectively. $g(x)$ is the hidden layer function and b_i is the threshold. Suppose there are N samples (x_i, t_i), $1 \le i \le N$, where $x_i = [x_{i1}, x_{i2}, \cdots, x_{im}]^{\mathrm{T}}$, $t_i = [t_{i1}, t_{i2}, \cdots, t_{in}]^{\mathrm{T}}$, the networking training model of ELM can be shown in Figure 4.

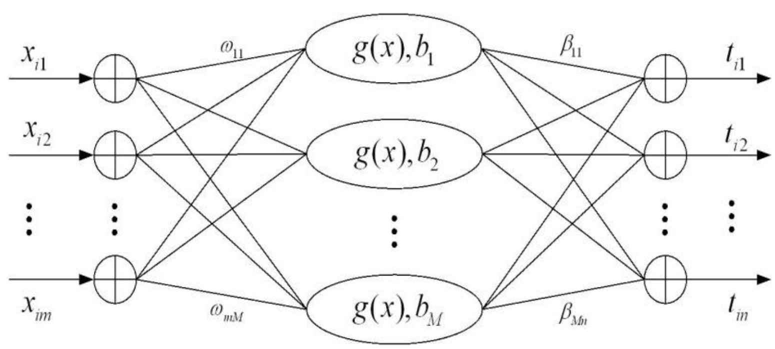


Figure 4. Network Training Model of ELM

The mathematical expression of the training model is as (10).

$$\sum_{i=1}^{M} \beta_i g(\omega_i x_i + b_i) = o_j,\ j = 1, 2, \cdots, N \tag{10}$$

Where $\omega_i = [\omega_{1i}, \omega_{2i}, \cdots, \omega_{mi}]^{\mathrm{T}}$, which denotes the input weights. $\beta_i = [\beta_{i1}, \beta_{i2}, \cdots, \beta_{in}]^{\mathrm{T}}$, which denotes the network output weights. $o_j = [o_{j1}, o_{j2}, \cdots, o_{jn}]^{\mathrm{T}}$, which denotes the network output vector.

The cost function of ELM can be expressed by (11).

$$E(S, \beta) = \sum_{j=1}^{N} \| o_j - t_j \| \tag{11}$$

Where $S=(\omega_i, b_i, i=1,2,\cdots,M)$, which contains the network input weights and the hidden layer thresholds. The ELM training process aims to find the optimal S and β to minimize the errors between the network output value and the actual value, that is $\min(E(S,\beta))$. The $\min(E(S,\beta))$ can be further expressed by (12).

$$\min(E(S,\beta))=\min_{\omega_i,b_i,\beta}\|H(\omega_1,\cdots,\omega_M,b_1,\cdots,b_M,x_1,\cdots,x_N)\beta-T\| \tag{12}$$

In (12), H is the hidden output matrix calculated by input samples. β is the output weights matrix and T is the objective value matrix. The H, β, T are defined as follows, respectively.

$$H(\omega_1,\cdots,\omega_M,b_1,\cdots,b_M,x_1,\cdots,x_N)=\begin{bmatrix} g(\omega_1 x_1+b_1) & \cdots & g(\omega_M x_1+b_M) \\ \vdots & & \vdots \\ g(\omega_1 x_N+b_1) & \cdots & g(\omega_M x_N+b_M) \end{bmatrix}_{N\times M} \tag{13}$$

$$\beta=\begin{bmatrix} \beta_1^{\mathrm{T}} \\ \vdots \\ \beta_M^{\mathrm{T}} \end{bmatrix}_{M\times n}, T=\begin{bmatrix} t_1^{\mathrm{T}} \\ \vdots \\ t_M^{\mathrm{T}} \end{bmatrix}_{N\times n} \tag{14}$$

The network training process of ELM can be treated as a nonlinear optimization problem, the objective function of which is expressed by (12). In case the hidden layer function is infinitely continuously derivable, the input weights and the hidden layer thresholds can be randomly assigned, and H would be a constant matrix. Therefore, the learning process of ELM can be equated with the calculation of the least square solution with minimum norm of the linear system $H\beta=T$. The calculation equation is as follows:

$$\hat{\beta}=H^{\dagger}T \tag{15}$$

Where $H^{\dagger}$ is the Moore-Penrose generalized inverse matrix of H. The network training can be realized by calculating $\hat{\beta}$. In the logical connectivity model, there are three inputs and one output. The inputs include the roadway length, vehicle number and vehicle speed. The output is the connectivity probability.

4. Simulations and Results

We compare the accuracy of the two proposed different network connectivity models for a set of parameters: roadway length, vehicle number and vehicle speed, which are shown in Table 1. Changing different values of the three parameters, 720 sets of data are collected through NS-2 simulation. These data constructs a four-dimensional data space (*p-lnv*). Part of the data is shown in Table 3.

Table 3. Part of the Simulation Data

l(m)	*n*	*v*(m/s)	*p*	*l*(m)	*n*	*v*(m/s)	*p*	*l*(m)	*n*	*v*(m/s)	*p*
800	10	5	0.3663	1400	30	5	0.5782	2000	50	5	0.454
800	10	10	0.2941	1400	30	10	0.5325	2000	50	10	0.5616
800	10	15	0.1746	1400	30	15	0.3057	2000	50	15	0.4060
800	10	20	0.1313	1400	30	20	0.1941	2000	50	20	0.2029
800	10	25	0.1148	1400	30	25	0.1607	2000	50	25	0.1917
800	10	30	0.1051	1400	30	30	0.1380	2000	50	30	0.1682
800	10	35	0.0943	1400	30	35	0.1302	2000	50	35	0.1468
800	10	40	0.0951	1400	30	40	0.1283	2000	50	40	0.1336
800	10	45	0.0782	1400	30	45	0.1192	2000	50	45	0.1305
800	10	50	0.0773	1400	30	50	0.1256	2000	50	50	0.1426
⋮	⋮	⋮	⋮	⋮	⋮	⋮	⋮				720×4

600 of the 720 sets of data are selected for nonlinear regression and ELM training, and others are used for testing. Making any two of the three parameters constant and the third one variable, we test the connectivity prediction results in three cases: 1) Case Ⅰ, the roadway length and vehicle number are constant, vehicle speed changes, 2) Case Ⅱ, the roadway length and vehicle speed are constant, vehicle number changes, 3) Case Ⅲ, the vehicle number and vehicle speed are constant, roadway length changes. Part of the prediction results are shown in Figure 5~7, respectively.

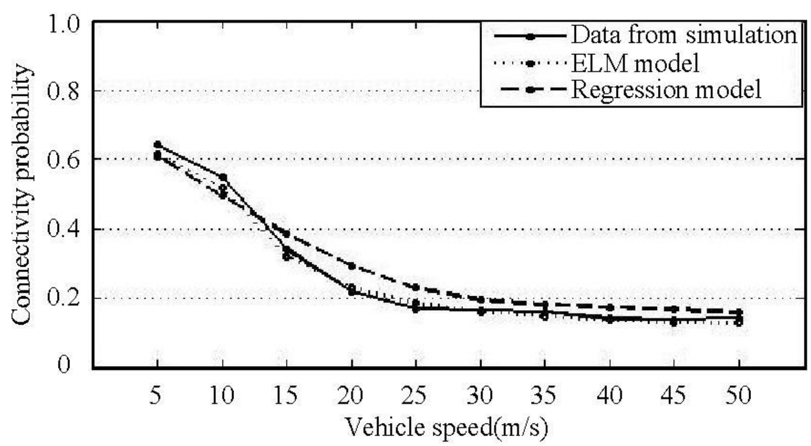


Figure 5. Case I: Roadway Length and Vehicle Number are Constant, Vehicle Speed Changes (*l*=1200m, *n*=25)

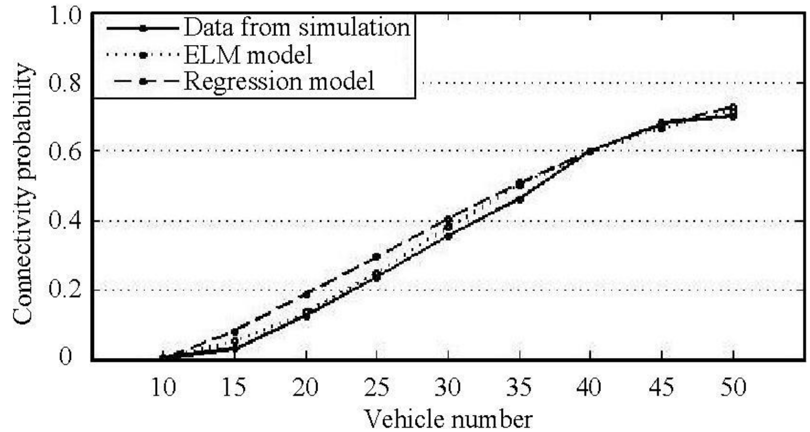


Figure 6. Case II: Roadway Length and Vehicle Speed are Constant, Vehicle Number Changes (*l*=1600m, *v*=5m/s)

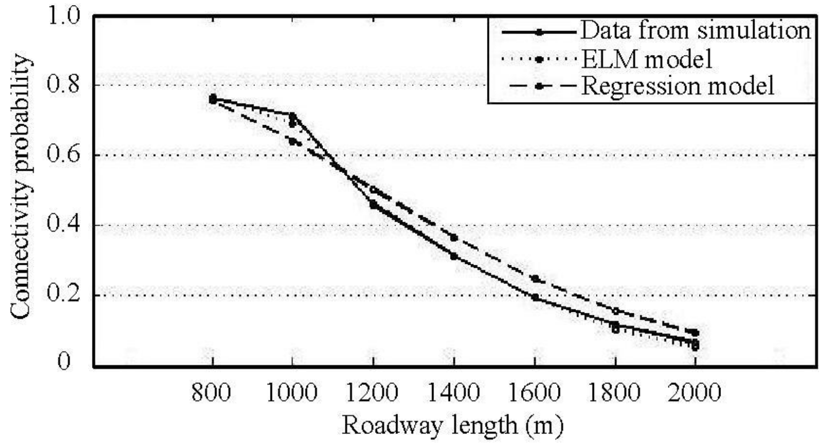


Figure 7. Case III: Vehicle Number and Vehicle Speed are Constant, Roadway Length Changes (*n*=35, *v*=20m/s)

As is shown in Figure 5-7, with different roadway lengths, vehicle speeds and vehicle numbers, the two connectivity models match the simulation value very well. In order to compare the accuracy of the two models, we use the root mean squared error (RMSE), which is expressed by (16), to analyze the predictions.

$$RMSE = \sqrt{\frac{1}{N}\sum_{k=1}^{N}[\hat{p}_k - f(l_k, n_k, v_k)]^2} \tag{16}$$

The smaller the value of RMSE is, the closer the prediction data approximates to simulation data, and the better performance the model is with. The RMSEs of all testing samples and the samples used in Figure 5-7 are calculated respectively. The results are shown in Table 4.

Table 4. RMSEs of Different Samples

Samples		All testing samples	Samples of Case Ⅰ	Samples of Case Ⅱ	Samples of Case Ⅲ
RMSE	Regression Model	0.0364	0.0384	0.0276	0.0411
	ELM model	0.0231	0.0149	0.0164	0.0104

From Table 4, we can find that the two models offer an extremely high accuracy (both the mean square errors of the two models are less than 5%). It implies that the two data mining methods are suitable for the VANET logical connectivity analysis and prediction. Furthermore, the ELM model is better than the regression model. The reason is that the regression method we used in our work is the basic least square method, which can smooth some characteristics of the training data (this inference can also be verified by curves shown in Figure 5-7), causing bigger errors.

5. Conclusions and Future Work

In this paper, taking the AODV protocol as an example, we study the logical connectivity of VANET by simulations. Firstly, we discuss the effects of different road traffic parameters on logical connectivity probability and select three main factors, the roadway length, vehicle number and vehicle speed. Furthermore, we propose two different connectivity prediction models based on nonlinear regression and ELM. Simulation results show that both of the two models are with high accuracy in predicting the connectivity under different road traffic status.

The future work includes two aspects. On one hand, to weaken the smoothing effect of regression function and improve the predicting accuracy, we will take the weight function and the Bicubic interpolation method into consideration in regression analysis. On the other hand, since AODV is the only protocol we analyze and model in this paper, future research also considers the logical connectivity of other routing protocols which are usually used in VANET, such as DSDV, GSR and GPSR.

References

[1] S. C. Lo, J. S. Gao and C. C. Tseng, “A water-wave broadcast scheme for emergency messages in VANET”, Wireless Personal Communications, vol. 71, no. 1, **(2013)**, pp. 217-241.

[2] A. G. EmadEddin, P. Evitm, G. Richard and Y. Y. Li, “Utilization of broadcast methods for detection of the road conditions in VANET”, 26th European Conference on Modeling and Simulations, Koblenz, Germany, **(2012)**.

[3] H. J. Chang and G. T. Park, “A study on traffic signal control at signalized intersections in vehicular ad hoc networks”, Ad Hoc Networks, vol. 11, no. 7, **(2013)**, pp. 2115-2124.

[4] B. X. Weis and A. Sandweg, “GuideWeb: a conceptually infrastructure-free vehicle navigation system”, IET Intelligent Transport Systems, vol. 6, no. 2, **(2012)**, pp. 139-144.

[5] N. P. Chandrasekharamenon and B. AnchareV, “Connectivity analysis of one-dimensional vehicular ad hoc networks in fading channels”, Journal on Wireless Communications and Networking, **(2012)**, vol. 2012.

[6] W. Xiong and Q. Q. Li, “Prerequisite to 1-connected vehicular ad hoc networks in highway scenarios”, Journal of Software,vol. 21, no. 11, **(2010)**, pp. 2906-2919. (in Chinese)

[7] N. Wisitpingphan, F. Bai, P. Mudalige and O. Tonguz, “On the routing problem in disconnected vehicular ad-hoc networks”, 26th IEEE international conference on computer communications. Piscataway, USA, **(2007)**, pp. 2291-2295.

[8] Z. J. Yan, H. Jiang, Z. Shen, Y. L. Chang and L. J. Huang, “k-connectivity analysis of one-dimensional linear VANETs”, IEEE Transactions on Vehicular Technology, vol. 61, no. 1, **(2012)**, pp. 426-433.

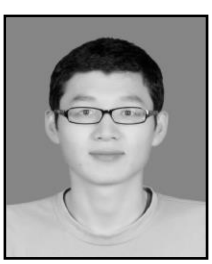
[9] P. C. Neelakantan and A. V. Babu, "Connectivity analysis of vehicular ad hoc networks from a physical layer perspective", Wireless Pers Commun, no. 71, **(2013)**, pp. 45-70.
[10] W. L. Jin and W. W. Recker, "An analytical model of multhop connectivity of inter-vehicle communication systems", IEEE Transactions on Wireless Communication Systems, vol. 9, no. 1, **(2010)**, pp. 106-112.
[11] Q. Yang, A. Lim, S. Li, J. Fang and P. Agrawal, "ACAR: adaptive connectivity aware routing for vehicular ad hoc networks in city scenarios", Mobile Netw Appl, no. 15, **(2010)**, pp. 36-60.
[12] C. Chen, X. B. Du, Q. Q. Pei and Y. N. Jin, "Connectivity analysis for free-flow traffic in VANETs: a statistical approach", International Journal of Distributed Sensor Networks, **(2013)**, vol. 2013.
[13] X. Jin, W. J. Su and Y. Wei, "A study of the VANET connectivity by percolation theory", 2011 IEEE Consumer Communications and Networking Conference, Las Vegas, USA, **(2011)**, pp. 85-89.
[14] G. B. Huang, Q. Y. Zhu and C. K. Siew, "Extreme learning machine: theory and applications", Neurocomputing, no. 70, **(2006)**, pp. 489-501.
[15] G. Y. Gao and G. P. Jiang, "Prediction of multivariable chaotic time series using optimized extreme learning machine", Acta Physica Sinica, vol. 61, no. 4, **(2012)**, pp. 37-45. (in Chinese)

Authors

Liu Haiqing, he is currently enrolled for joint courses for master and doctor degrees in School of Control Science and Engineering, Shandong University. His research interests include intelligent transportation system, vehicular ad hoc network and cooperative vehicle infrastructure system.

Yang Licai, he is currently a professor and doctoral supervisor at School of Control Science and Engineering in Shandong University. His research interests include artificial intelligence and intelligent control, control theory and applications, intelligent transportation systems, and vehicular ad hoc networks.

Ding Sijing, she is a graduate student of School of Control Science and Engineering, Shandong University, majoring in Control Science and Engineering. Her research interests include intelligent transportation system and vehicular ad hoc network.



Shi Yun-feng, he is currently a Ph.D. candidate in School of Control Science and Engineering, Shandong University. His research interests include vehicular ad hoc network and cooperative vehicle infrastructure systems.