Research Article

Reinforcement Learning-Based Multiple Constraint Electric Vehicle Charging Service Scheduling

Yongguang Liu,[1] Wei Chen [ID],[2] and Zhu Huang[2]

[1]*Henan Xj Metering Co., Ltd, Xuchang 450061, Henan, China*
[2]*Hefei University of Technology, Hefei City 230041, Anhui Province, China*

Correspondence should be addressed to Wei Chen; 2020110419@mail.hfut.edu.cn

The popularization of electric vehicles faces problems such as difficulty in charging, difficulty in selecting fast charging locations, and comprehensive consideration of multiple factors and vehicle interactions. With the increasingly mature application of navigation technology in vehicle-road coordination and other aspects, the proposal of an optimal dynamic charging method for electric fleets based on adaptive learning makes it possible for edge computing to process electric fleets to effectively execute the optimal route charging plan. We propose a method of electric vehicle charging service scheduling based on reinforcement learning. First, an intelligent transportation system is proposed, and on this basis a framework for the interaction between fast charging stations and electric vehicles is established. Subsequently, a dynamic travel time model for traffic sections was established. Based on the habits of electric vehicle owners, an electric vehicle charging navigation model and a reinforcement learning reward model were proposed. Finally, an electric vehicle charging navigation scheduling method is proposed to optimize the service resources of the fast charging stations in the area. The simulation results show that the method balances the charging load between stations, can effectively improve the charging efficiency of electric vehicles, and increases user satisfaction.

1. Introduction

With the extensive development of electric vehicles in various countries around the world, the number of electric vehicles is increasing, and problems such as difficulty in charging electric vehicles, serious line losses, voltage drops, charging safety, and severe peaks are expected [1–3]. Electric vehicle charging and charging path planning should receive more attention. For electric vehicles whose driving time is longer than the nondriving time, fast charging is an important power supplement method [4, 5]. The disorderly charging of electric vehicles would not only cause congestion of fast charging stations, which increases the burden on the regional grid, but also result in concentrated charging times causing problems such as transformer overload and increased peak-to-valley difference, which is not conducive to the safe operation of the distribution network [6, 7]. Therefore, reasonable guidance and charging scheduling for vehicles with fast charging needs are beneficial to the

alleviation of the burden on the regional grid while meeting the charging needs [8, 9].

In response to the above problems, scholars at home and abroad have conducted some research. In [10], we studied the uniform charging node in [11] and extended it to the nonuniform charging node in [12] by solving the mixed integer nonlinear programming problem (MINLP) of the single vehicle. The remaining energy of the vehicle on each node is expressed as a dynamic programming (DP) problem for a single electric vehicle path problem, and a DP-based algorithm is provided to determine the optimal path and charging strategy of the electric vehicle subflow level. In [13], we proposed a distributed electric vehicle path selection system based on the distributed ant colony algorithm (ACA). The distributed architecture minimizes the total travel of electric vehicles to the destination by proposing a set of nearest fast charging stations. In [14], we proposed an improved Dijkstra method to solve the multiobjective optimization problem and obtained a multiobjective

optimization function including travel time, fast charging station number of vehicles, and charging load, thereby optimizing electric vehicle charging path planning and alleviating fast charging stations. The lack of surrounding traffic congestion reduces waiting time and improves the availability of charging facilities.

The above literature has its own characteristics regarding charging route navigation and charging scheduling, but when studying electric vehicle charging route navigation, it only focuses on the economic benefits and waiting time of the vehicle and ignores the impact of fast charging station loads when charging large-scale electric vehicles. Most charging scheduling uses a fixed strategy while ignoring the influence of various factors, such as the increase in the number of electric vehicles and user habits, on electric vehicle charging scheduling for different time periods.

In this context, we propose an electric vehicle charging service scheduling method based on reinforcement learning to meet the needs of electric vehicle owners. The structure of the paper is as follows. In Section 2, we propose a fast charging station and electric vehicle system framework and use this framework to study electric vehicle charging navigation. In Section 3, we establish a dynamic travel time model for traffic sections and propose an electric vehicle charging navigation model. In Section 4, incorporating reinforcement learning, we further propose an electric vehicle charging navigation scheduling method to rationally optimize the service resources of each fast charging station in the area. In Section 5, we use a certain city as a model and compare the simulation results of the proposed method with those of the traditional electric vehicle charging navigation method to demonstrate the superiority of this method. Conclusions and further research directions are outlined in Section 6.

2. Fast Charging Station and Electric Vehicle System Framework

With the gradual development and application of 4G and 5G communications, the applications of various technologies for navigation and vehicle-road collaboration have become increasingly mature [15, 16]. At the same time, edge computing technology also provides technical guarantees for fast response and low error rate operating environments. The computational burden of the central scheduling node is transferred to the user edge side, which greatly increases the processing efficiency and enables electric vehicles and fast charging stations to share information and synchronize processing [17].

Currently, electric vehicles can share information with fast charging stations and other systems through the Internet, upload the status and location of electric vehicles in real time, and navigate in real time based on the location of electric vehicles [18, 19]. Moreover, a variety of optimal dynamic charging methods for electric fleets based on adaptive learning have been proposed, and the results show that this method can basically achieve the optimal solution. On this basis, the optimal route charging schedule can be effectively carried out for the electric fleet of efficient and

dynamic transportation systems. Inspired by the above research, this paper proposes a guidance system structure for electric vehicles and fast charging stations. The structure of the guidance system for electric vehicles and fast charging stations in this article is shown in Figure 1. With the Internet platform as the center, the system dynamically updates intersection information and provides dynamic charging and navigation strategies for electric vehicles by referring to road condition information and fast charging station information. Navigation combines the road condition information and the waiting time of each fast charging station and chooses the fast charging station with the highest overall efficiency for itself to charge. The fast charging station itself further charges the electric vehicle according to various factors, such as weather, energy supply and demand, and user habits. At intervals, the traffic information and fast charging station information are refreshed according to the above selection, and the charging navigation strategy is provided again.

3. Preliminary Model Establishment

This section first proposes a dynamic travel time model for traffic sections and, on this basis, establishes a charging navigation model that considers distance, time, and economic benefits for a single electric vehicle.

3.1. Dynamic Travel Time Model of Traffic Section. The dynamic path selection model for electric vehicles in this paper is based on the dynamic travel time model of the road segment. First, the movement of the vehicle in the road segment is described by the cumulative number of vehicles $M(a, t)$, which represents the sum of the number of vehicles passing observation point a before time t. According to the definition of flow and density, the traffic flow $\sigma(a, t)$ and traffic density $\rho(a, t)$ are as follows:

$$\sigma(a, t) = \lim_{t \to t_0} \left(\frac{M(a, t) - M(a, t_0)}{t - t_0} \right) = \frac{\partial M(a, t)}{\partial t},$$

$$\rho(a, t) = \lim_{a \to a_0} \left(\frac{M(a_0, t) - M(a, t)}{a - a_0} \right) = \frac{\partial M(a, t)}{\partial a},$$

(1)

where $M(a, t_0)$ and $M(a_0, t)$ are the number of vehicles at position a at time t_0 and the number of vehicles at position a_0 at time t, respectively.

According to the traffic volume and traffic density, the traffic velocity $v(a, t)$ can be obtained as follows:

$$v(a, t) = \frac{\sigma(a, t)}{\rho(a, t)} = \frac{\partial M(a, t)}{\partial t} \cdot \frac{\partial a}{\partial M(a, t)}.$$

(2)

Assuming that the vehicles on the road section are evenly distributed in the road section, the traffic density $\rho_i(t)$ of road section i is as follows:

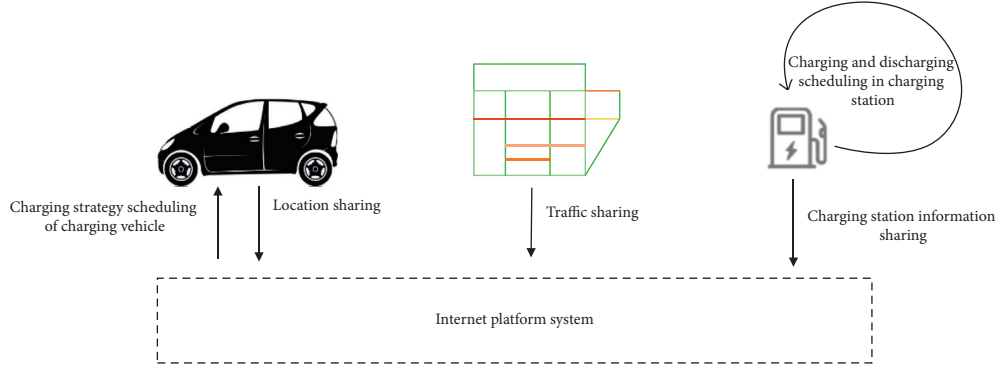$$\rho_i(t) = \frac{M(a_i^0, t) - M(a_i^L, t)}{L_i \cdot n_i},$$

(3)

FIGURE 1: Electric vehicle and charging station guidance system structure.

where a_i^0 and a_i^L are the entrance and exit positions of road section i, respectively; n_i is the number of vehicles that can be accommodated per unit length in road section i; and L_i is the length of road section i.

According to the above formula, the vehicle speed $v_i(t)$ on road section i can be expressed as follows [20]:

$$v_i(t) = \begin{cases} v_{i,\text{free}}, & \rho_i(t) < \rho_{i,\min}, \\ v_{i,\min} + \left(v_{i,\text{free}} - v_{i,\min}\right)\left[1 - \left(\dfrac{\rho_i(t) - \rho_{i,\min}}{\rho_{i,\max} - \rho_{i,\min}}\right)^\alpha\right]^\beta & \rho_{i,\min} \leq \rho_i(t) \leq \rho_{i,\max}, \\ v_{i,\min}, & \rho_i(t) > \rho_{i,\max}, \end{cases} \tag{4}$$

where $v_{i,\text{free}}$ is the free flow velocity of section i; $\rho_{i,\max}$ and $\rho_{i,\min}$ are the maximum density and minimum density on section i, respectively; $v_{i,\min}$ is the minimum vehicle speed; and α and β are system model parameters.

It can be concluded that the passing time T_i of road section i is expressed as follows:

$$T_i = \frac{L_i}{v_i(t)} = \frac{L_i}{(\partial M(a,t)/\partial t) \cdot (\partial a/\partial M(a,t))}. \tag{5}$$

If the road congestion signal is received halfway, the system changes the route to reduce the delay time. The subjective probability of the owner changing road section i to road section i' is $P_{i \longrightarrow i'}$:

$$P_{i \longrightarrow i'} = \frac{\int_0^{T_i} e^{-\left(\left(T_{Max} - T_{i'}\right)^2/2(\eta)^2\right)} \mathrm{d}\left(T_{\text{Max}}\right)}{\sqrt{2\pi}\eta}, \tag{6}$$

where T_i is the travel time of section i in the route; $T_{i'}$ is the travel time of section i' in the route; T_{Max} is the maximum travel time; and η is a subjective coefficient.

Therefore, the length of the driving section can be approximated by subjective probability as d_i:

$$\begin{aligned} d_i &= \left(1 - P_{i \longrightarrow i'}\right)L_i + P_{i \longrightarrow i'} L_{i'} \\ &= \left[1 - \frac{\int_0^{T_i} e^{-\left(\left(T_{Max} - T_{i'}\right)^2/2(\eta)^2\right)} \mathrm{d}\left(T_{\text{Max}}\right)}{\sqrt{2\pi}\,\eta}\right] \cdot L_i + \frac{\int_0^{T_i} e^{-\left(\left(T_{Max} - T_{i'}\right)^2/2(\eta)^2\right)} \mathrm{d}\left(T_{\text{Max}}\right)}{\sqrt{2\pi}\eta} \cdot L_{i'}, \end{aligned} \tag{7}$$

where $L_{i'}$ is the length of road section i'.

3.2. Electric Vehicle Charging Navigation Model. Electric vehicles need to be charged frequently during use, so there will be demand for fast charging. According to the charging needs of different vehicles, implementing different navigation schemes can effectively improve the response speed of the vehicle. This section comprehensively considers the driving distance required to reach a fast charging station, the total time of driving and charging, and the charging economy to establish a charging navigation model.

For electric vehicle owners with high total driving distance requirements, this article considers the principle that the direction of the fast charging station is the same as the destination direction when all vehicles are connected to the Internet. It is proposed that the sum of the shortest distance $\min D$ from the starting point O of the vehicle to the fast charging station S and from the fast charging station S to the destination D is expressed as follows:

$$\min D = \sum_{a=1}^{m} \sum_{b=1, b \neq a}^{m} d_{ab,OS} \alpha_{ab} + \sum_{a=1}^{m} \sum_{b=1, b \neq a}^{m} d_{ab,S\,D} \alpha_{ab}, \quad (8)$$

where a and b are path nodes; m is the total number of path nodes; $d_{ab,OS}$ and $d_{ab,S\,D}$ are the length of the road section from the starting point O to the fast charging station S and from the fast charging station S to the destination D with a and b as the end nodes; and α_{ab} is a variable that equals 1 for the road section with a and b as the end nodes and equals 0 otherwise.

For electric vehicle owners with high total time requirements, this article proposes the shortest total charging time as the goal to optimize the charging path:

$$\min T = T_D + T_C + T_Q. \quad (9)$$

The specific solutions of T_D and T_C are as follows:

$$T_D = \sum_{i=1}^{i=m} \frac{d_i}{v_i(t)}$$

$$= \sum_{i=1}^{i=m} \frac{(1 - P_{i \longrightarrow i'})L_i + P_{i \longrightarrow i'} L_{i'}}{v_i(t)}$$

$$= \sum_{i=1}^{i=m} \frac{(1 - P_{i \longrightarrow i'})L_i + P_{i \longrightarrow i'} L_{i'}}{(\partial M(a,t)/\partial t) \cdot (\partial a/\partial M(a,t))}, \quad (10)$$

$$T_C = \frac{Q_{Ex} - Q_{Re}}{P\theta}$$

$$= \frac{Q_{Ex} - \left(C_{car} \cdot C_{carINI} - Q \int_0^{T_D} v_i(t) dt\right)}{P \cdot \theta},$$

where T_D is the travel time to the fast charging station; T_Q is the waiting time in the fast charging station, which is determined by the number of vehicles; T_C is the charging time; Q_{Ex} is the expected voltage at the end of charging, which is set to 95% of the full charge; Q_{Re} is the remaining power to the fast charging station; P is the charger power; θ is the charging efficiency; C_{car} is the electric vehicle battery capacity; C_{carINI} is the initial state of charge of the electric vehicle; and Q is the electric energy consumed by the electric vehicle per kilometer.

For electric vehicle owners with high cost requirements, this article proposes the minimum cost as the goal to optimize the charging path:

$$\min M = M_D + M_S, \quad (11)$$

where M_D is the electricity cost consumed on the charging path and M_s is the cost consumed by the fast charging station.

4. Electric Vehicle Charging Navigation Scheduling Strategy Based on Reinforcement Learning

The goal of the reinforcement learning algorithm is to find an optimal strategy based on the Markov decision process to maximize the expected cumulative return. In this section, the driving distance of the electric vehicle, the total driving and charging time, and the charging economy are optimized in parallel to provide the electric vehicle owner with the best electric vehicle charging navigation scheduling strategy [21, 22].

4.1. Strategy Gradient Algorithm. The basic principle of reinforcement learning is to learn from exploratory experiments and obtain action strategies to achieve established goals. The learning subject is the agent; the object interacting with the agent is the environment. Reinforcement learning is an abstraction of goal-oriented interactive learning problems. In a certain environment state, the agent takes action, and the environment responds to the agent's actions, presents the new environment state to the agent, and feeds a certain reward back to the agent. The agent and the environment continue to interact to achieve the ultimate goal of maximizing returns.

The interaction process between the agent and the environment can be described by a time series: in a certain period t, the agent takes a certain action a according to the current environment state s_t^n; in the next period $t + 1$, due to the agent's action a_t^n, the environment state changes from s_t^n to s_{t+1}^n, and the agent is rewarded with $r(t)^n$. In each time period, the probability distribution of all actions that the agent can take in the current environment state is called the agent's strategy π. The agent continuously changes its strategy through interaction and finally achieves the goal of maximizing rewards.

The reinforcement learning problem satisfies the Markov characteristic; that is, the state of the next period is only related to the state s_t^n of the current period and has nothing to do with the state s_{t+1}^n of the previous period. The policy-based method is used to express a policy. Assuming that the strategy of electric vehicle charging and navigation control consists of a t-step decision, the agent obtains n corresponding training trajectories τ_n by interacting with the environment as follows:

$$\tau_n = [s_1^n, r(1)^n, a_1^n, s_2^n, r(2)^n, a_2^n, \ldots, s_t^n, r(t)^n, a_t^n], \quad (12)$$

where a_t^n represents the action determined at time t during the n training, s_t^n represents the state after action a during the n training, and $r(t)^n$ represents the reward obtained after action a during the n training. The expected return reward R_θ for all stored trajectories is as follows:

$$R_\theta = \sum_{n=1} R(\tau_n) \bullet p_\theta(\tau_n)$$

$$= \sum_{n=1} R(\tau_n) \bullet \prod_{t=1}^{T} [p(a_t^n|s_t^n, \theta) \cdot p(r(t+1)^n, s_{t+1}^n|s_t^n, a_t^n)], \quad (13)$$

where $R(\tau_n) = \sum_{i=0}^{t} r(t)^n$ is the reward value of trajectory τ_n, $p_\theta(\tau_n)$ is the probability of trajectory τ_n, $p(r(t+1)^n, U_{t+1}^n, f_{t+1}^n|U_t^n, f_t^n, a_t^n)$ is the probability of

$[r(t+1)^n, s_{t+1}^n]$ in state $[s_t^n, a_t^n]$, and $p(a_t^n|s_t^n, \theta)$ is the probability of selecting actions a_t^n according to input and output strategy $\pi_n(\theta)$ in state s_t^n.

Therefore, reinforcement learning can be expressed as solving the maximum expected return reward R_θ. To realize the strategy, the partial derivative of the parameter set θ is obtained to obtain the optimized strategy function ∇R_θ as follows:

$$\nabla R_\theta = \nabla \sum_{n=1} R(\tau_n) \bullet p_\theta(\tau_n)$$

$$= \sum_{n=1} R(\tau_n) \bullet \nabla p_\theta(\tau_n)$$

$$= \sum_{n=1} R(\tau_n) \bullet p_\theta(\tau_n) \frac{\nabla p_\theta(\tau_n)}{p_\theta(\tau_n)}$$

$$= \sum_{n=1} R(\tau_n) \bullet p_\theta(\tau_n) \bullet \nabla \log p_\theta(\tau_n)$$

$$= \frac{1}{N} \sum_{n=1}^{N} R(\tau_n) \bullet \nabla \log p_\theta(\tau_n)$$

$$= \frac{1}{N} \sum_{n=1}^{N} R(\tau_n) \bullet \nabla \log \left\{ \prod_{t=1}^{T} [p(a_t^n|s_t^n, \theta) \bullet p(r(t)^n, s_{t+1}^n|s_t^n, a_t^n)] \right\} \quad (14)$$

$$= \frac{1}{N} \sum_{n=1}^{N} R(\tau_n) \bullet \nabla \left\{ \sum_{t=1}^{T} [\log p(a_t^n|s_t^n, \theta) + p(r(t)^n, s_{t+1}^n|s_t^n, a_t^n)] \right\}$$

$$= \frac{1}{N} \sum_{n=1}^{N} R(\tau_n) \bullet \nabla \sum_{t=1}^{T} \log p(a_t^n|s_t^n, \theta)$$

$$= \frac{1}{N} \sum_{n=1}^{N} \sum_{t=1}^{T} R(\tau_n) \bullet \nabla \log p(a_t^n|s_t^n, \theta).$$

The reinforcement learning policy gradient algorithm is equivalent to solving a partial derivative problem. If the parameter set θ is updated in the positive direction, that is, the reward increases, the probability of trajectory τ_n will increase, and vice versa. The pseudocode of the policy gradient Algorithm 1 is given below.

4.2. Action Selection.
Taking the vehicle travel path as an example, the control parameter is a_t^n, and the vehicle has 3 possible actions at each intersection. The value range is $a_t^n = [0, 2]$. 0 means going forward, 1 means turning left, and 2 means turning right.

4.3. Environmental Status.
When an electric vehicle performs an action at an intersection and acts on the environment, the state value corresponding to the environmental feedback is s_t^n, that is, $s_t^n = [D_t^n, T_t^n, M_t^n]$, and the dimension is 3. Among them, D_t^n, T_t^n, and M_t^n correspond to the distance, time, and cost, respectively, after the current action is executed. After obtaining the environmental state value, the corresponding reward value is calculated, and at the same time, the environment will move to the next state.

4.4. Reward Function Design.
The reward function is designed as follows:

$$\text{reward} = -a \cdot (D_t^n - \min D)^2 - b \cdot (T_t^n - \min T)^2 - c \cdot (M_t^n - \min M)^2, \quad (15)$$

where reward represents the reward value obtained by the action performed by the electric vehicle at each time node, that is, the quality of the current trajectory τ_n action.

(1) In the neural network, initialize the parameter set θ randomly and initialize $n = 1$.
(2) Initialize $t = 0$, randomly initialize action a_t^n and output state s_{t+1}^n, calculate local reward $r(t+1)^n$, and then add the trajectory generated by the $t + 1$ action to the stored trajectory τ_n of the n training.
(3) Input state s_{t+1}^n to the neural network and select a random action a_{t+1}^n.
(4) After the simulation environment executes action a_{t+1}^n, obtains the output state s_{t+2}^n, and calculates the local reward $r(t+2)^n$, the trajectory generated by the $t + 2$ action is added to the stored trajectory τ_n of the n training.
(5) Judge whether $R(\tau_n) = \sum_{i=1}^{i=t} r(t)^n > R$ is true; if it is true, go to step 6; otherwise, assign $t + 1$ to t and go to step 3, where i is the variable to be accumulated and R is the expected value of the total reward for a single trajectory.
(6) Calculate the strategy optimization strategy function ∇R_θ.
(7) Assign $n + 1$ to n, update the parameter set θ in strategy $\pi_n(\theta)$ to $\theta + \gamma * \nabla R_\theta$, and judge whether $n \leq N$ is true; if so, go to step 2; otherwise, the reinforcement learning training process is over; save the updated parameter set as the most optimal parameter set θ^* and the optimal strategy π; N is the maximum number of trajectories

ALGORITHM 1: Policy gradient algorithm

Among them, a, b, and c are the weighting coefficients: when the owner only cares about the distance, a equals 1, and the rest equal 0; when the owner only cares about the total time, b equals 1, and the rest equal 0; when the owner only cares about the cost, c equals 1, and the rest equal 0; and if the owner chooses to focus on all three variables, set $a + b + c = 1$, and assign values according to the proportion.

4.5. Controller. For electric vehicle charging navigation, a scheduling algorithm based on the policy gradient algorithm is proposed according to the personal habits of different electric vehicle owners. By observing the information to select a behavior directly for back propagation and using rewards to directly enhance and weaken the possibility of selection behavior, the probability of selecting good behavior will increase next time, and bad behavior will be weakened next time.

A three-layer wavelet neural network is used. The wavelet neural network is a multilayer feedforward neural network trained according to error back propagation [23]. This article uses a three-layer neural network, that is, one output layer, one input layer, and one hidden layer, as shown in Figure 2. The state is set as the input layer of the neural network. Its dimension is 3; the hidden layer of the neural network has 20 neurons; and the output layer contains 3 neurons, corresponding to 3 output actions.

The connection weights and bias terms between the input layer and the hidden layer and between the hidden layer and the output layer are represented by a parameter set of θ. The input and output strategies of the n training wavelet neural network of the strategy body are defined as $\pi_n(\theta)$.

The activation function of the connection between the input layer and the hidden layer is $\mathrm{Tan}\,h$, and its function formula is as follows:

$$f(z) = \mathrm{Tan}\,h(z) = \frac{e^z - e^{-z}}{e^z + e^{-z}}. \tag{16}$$

The activation function connecting the hidden layer and the output layer is a wavelet basis function, and its function formula is as follows:

$$\varsigma(z) = e^{-\left(z^2/z\right)} \cos(1.75z). \tag{17}$$

According to the pseudocode of the algorithm, the specific training process can be obtained as shown in Figure 3.

5. Simulation Results and Discussion

Taking the city in Figure 4 as a model, the city includes 21 nodes, 32 road sections, and 4 fast charging stations. The number marked on the road section represents the length of the road section in km. Fast charging stations are located at nodes 9, 12, 14, and 19. For electric vehicles, the battery capacity is 90 kW·h, the cruising range is 400 km, and the fast charging station power is 350 kW. When the electric vehicle leaves the fast charging station, C_{carINI} is 90%; the training parameters are as follows: the number of training rounds is 1900, and the learning coefficient is 0.95. The discount rate is 0.95.

The vehicle randomly sets the initial position and target position (on 21 nodes) and randomly sets the remaining power (not higher than 30%). According to the distance selected by the user, the total time consumed, and the cost as the reward value, the vehicle is trained from the initial position to the fast charging station to charge and from the fast charging station to the target location. After the training is completed, the final reward changes are shown in Figure 5.

Figure 5 shows that as the number of training sessions increases, the training reward gradually increases. After 600 training sessions, the curve shows an oscillating trend, and the reward oscillates around 190. In the subsequent training, the reward is basically stable. Save the neural network model obtained from the last training parameter.

The 08:00 traffic flow distribution obtained through urban traffic simulation is shown in Figure 6. The green line represents smooth traffic, orange represents traffic congestion, and red represents heavy traffic congestion. For the traffic flow shown in Figure 6, the saved reinforcement learning model is used to obtain the station selection probability of the electric vehicle when each network node starts, as shown in Figure 7. It can be concluded that under the premise of considering congestion, the trained reinforcement learning model can effectively select fast charging
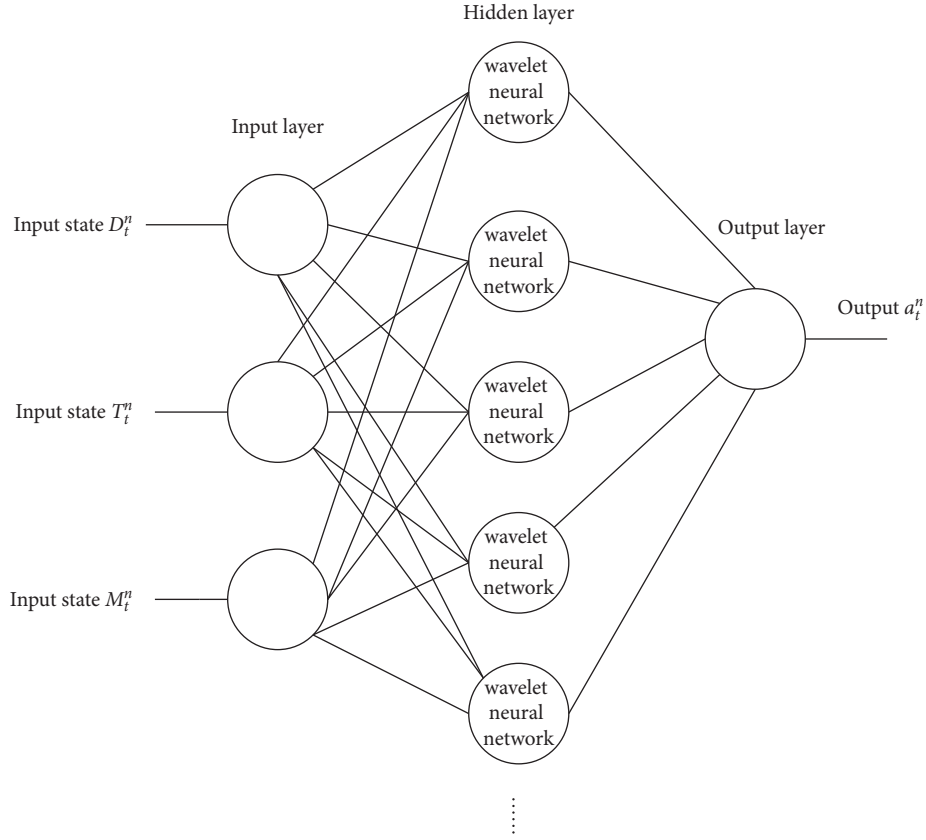
FIGURE 2: Structure diagram of wavelet neural network.

stations corresponding to shorter distances according to the target node.

Now, take an electric vehicle starting at node 13 and ending at node 2 as an example to analyze its dynamic station selection strategy. Consider the distance, total time, and cost required for the owner to obtain charging navigation during driving, as shown in Table 1.

Plan 1 takes the minimum distance as the goal and chooses fast charging station No. 9, and the travel route is shown as the solid line in Figure 8. Plan 2 takes the minimum time as the goal and chooses fast charging station No. 14, and the travel route is shown by the dashed line in Figure 8. Plan 3 takes the minimum cost as the goal and chooses No. 12 fast charging station, and the travel route is shown as the crossed line in Figure 8.

Multiple routes were selected for testing, and methods from [10, 13] and the charging navigation method proposed in this paper were compared. The performance comparisons under the comprehensive requirements of the research vehicle owners are shown in Figure 9.

The first graph in Figure 9 shows the change trend of the average distance with the increase in the number of test routes under the premise of considering the comprehensive performance required by the user. In this graph, the comparison between the method in this paper and the methods in the other two references is shown. With the increase in the number of routes, the average travel distance of the three methods fluctuated and finally stabilized in the vicinity of

17 km. In this process, the total distance predicted by the three methods is basically the same. The second graph in Figure 9 shows the trend of the total time as the number of test routes increases. As the number of routes increases, the total time of the method in this paper steadily decreases, and finally the time is reduced to 0.7 h, while for the other two methods, the total time consumed curve presents an oscillating situation, and the time consumed is unstable and greater than that for the method in this paper. It can be concluded from the curve that the method in this paper has the least total time consumption. The third graph in Figure 9 shows the trend of the total cost as the number of test routes increases. With the increase in the number of routes, the total cost of the method in this paper first increases, then gradually decreases, and finally stabilizes at approximately 30 yuan. For the method from [10], the total cost of the method was initially lower than that of the method in this paper. With the increase in the number of test routes, the cost began to increase and eventually was significantly higher than that of the method in this paper. The cost for the method in [13] remained higher than the cost for the method in this paper after initially oscillating lower. It can be concluded that, under the comprehensive performance requirements, the total distances of the three methods are basically the same. On this basis, with the increase in the number of route tests, the method in this paper has the least total time and cost, which indicates the superiority of the method in this paper.
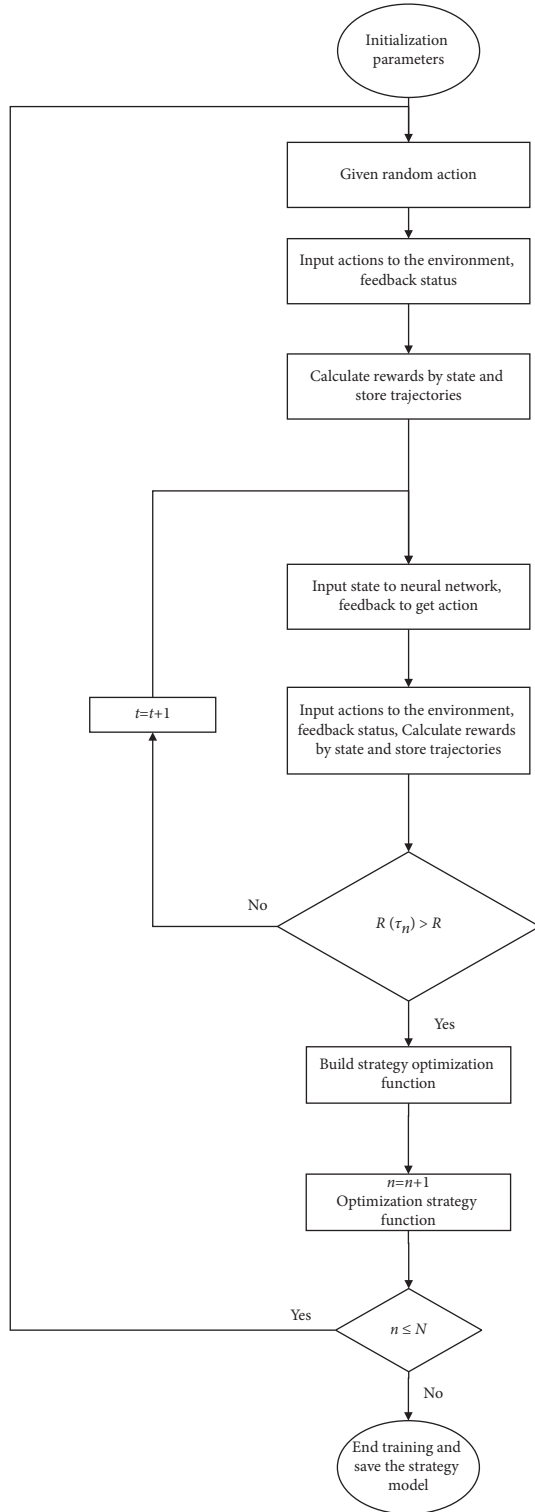
FIGURE 3: Flowchart of training steps.

FIGURE 4: City model chart.

FIGURE 5: Training result chart.

FIGURE 6: Traffic flowchart.

In the case of the same time, initial point, and destination, we compare user satisfaction under the electric vehicle charging navigation strategy in [10, 13, 24–26]. The
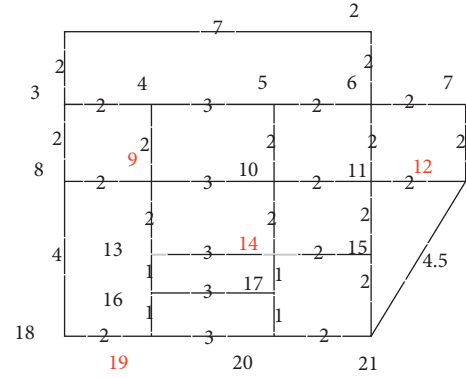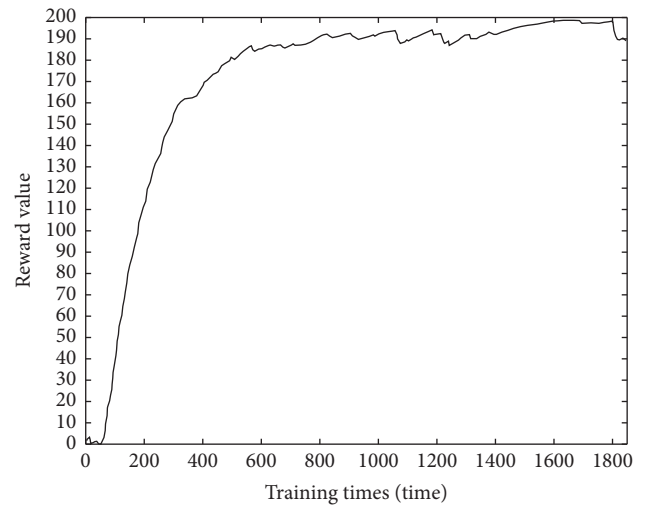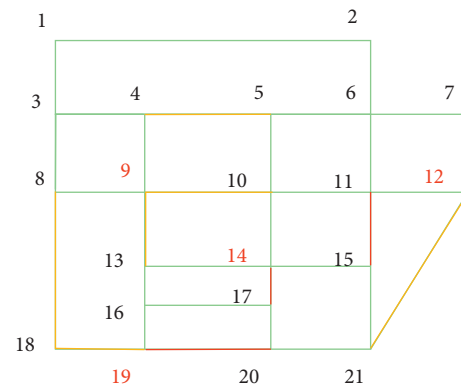
user satisfaction from testing electric vehicles using these methods is shown in Table 2. It can be seen in the table that, with the increase of test lines, user satisfaction under this
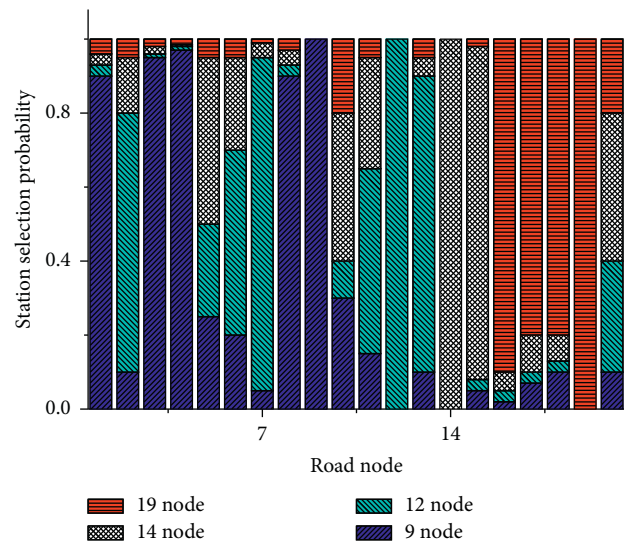
FIGURE 7: Probability of selecting a fast charging station node.

TABLE 1: Scheme comparison.

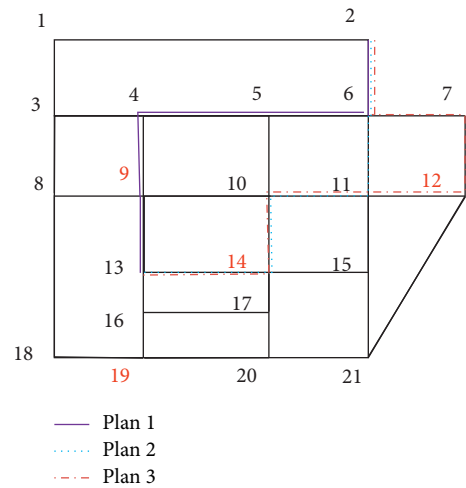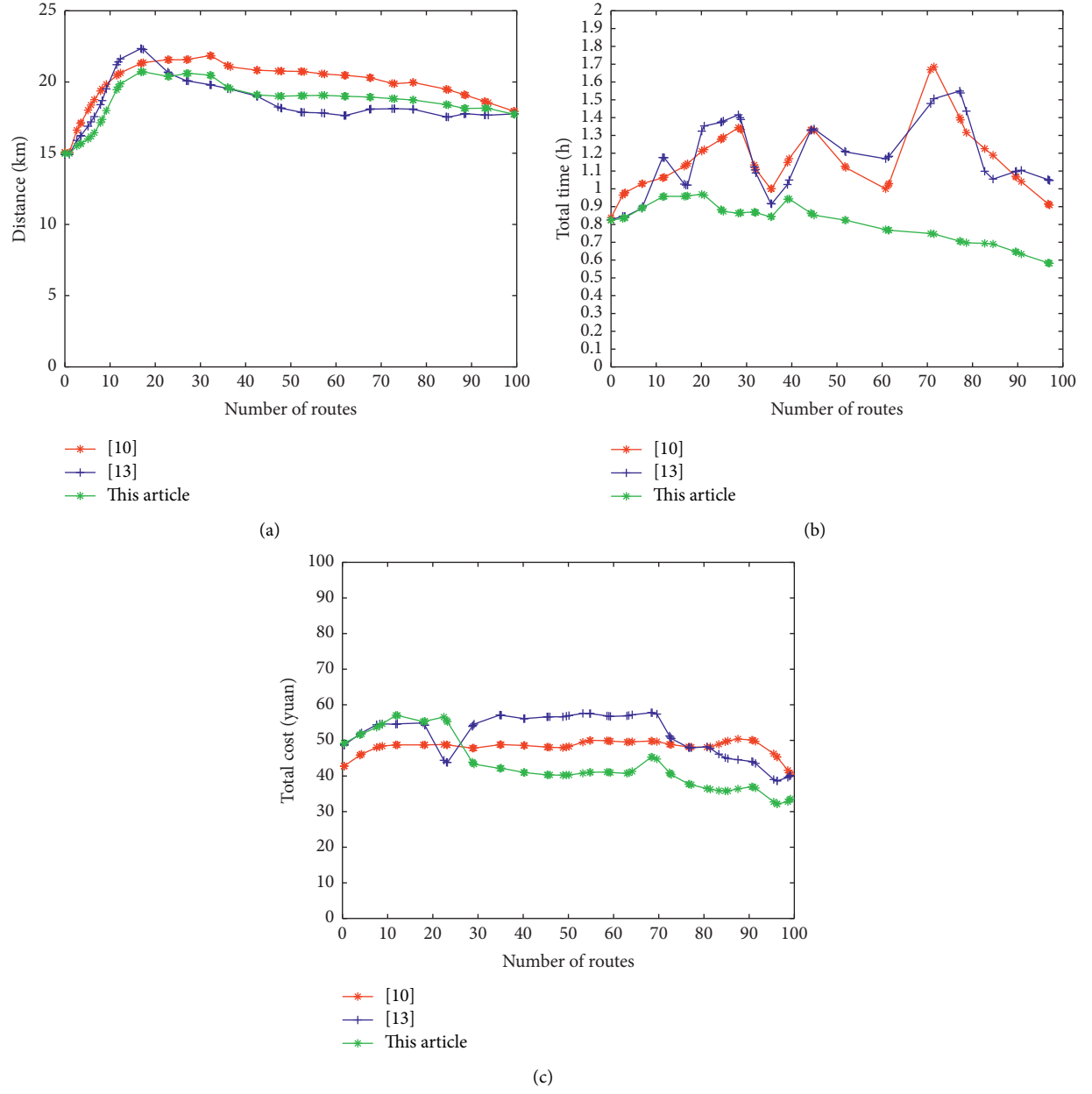| Plan	Distance (km)	Total time (h)	Expenses (yuan)	Station selection
1	11	1.5	76	9
2	11	1	84	14
3	15	1.8	60	12

FIGURE 8: Route selection chart.

(a)

(b)

(c)

Figure 9: Distance, time, and cost comparison chart.

Table 2: User satisfaction for electric vehicle charging navigation strategies.

Number of routes	This article (%)	[10] (%)	[13] (%)	[24] (%)	[25] (%)	[26] (%)
1	100	100	100	100	100	100
10	100	95	80	90	90	70
100	95	80	70	90	80	75
1000	90	75	75	85	75	55

method far exceeds other methods. It has been demonstrated that the method in the article can effectively meet the charging and navigation needs of users.

6. Conclusions

We propose an electric vehicle charging service scheduling method based on reinforcement learning to meet the needs of electric vehicle owners. First, based on an intelligent transportation system, a framework for the interaction between fast charging stations and electric vehicles is proposed. Subsequently, the dynamic travel time model of the traffic section was established, and the electric vehicle charging navigation model was proposed. Finally, combined with reinforcement learning, the electric vehicle charging navigation scheduling method is further proposed to rationally optimize the service resources of each fast charging station in the area. The results show that, compared with the existing methods, the algorithm and model proposed in this paper can effectively optimize electric vehicle charging and navigation scheduling based on the needs of the vehicle owner and can meet the various needs of the vehicle owner.

Data Availability

The MATLAB simulation data used to support the findings of this study are currently under embargo while the research findings are commercialized. Requests for data, 12 months after publication of this article, will be considered by the corresponding author.

Conflicts of Interest

The authors declare that there are no conflicts of interest regarding the publication of this study.

References

[1] X. Sun, Z. Li, and X. Wang, "Technology development of electric vehicles: a review," *Energies*, vol. 90, 2020.

[2] A. V. Prakash and S. Parameswaran, "Smart parking and charging management of electric vehicles in public parking space," in *Proceedings of the 2018 Second International Conference on Intelligent Computing and Control Systems (ICICCS)*, Madurai, India, 2018.

[3] S. H. Cutcliffe and D. A. Kirsch, "The electric vehicle and the burden of history," *History*, vol. 6, no. 2, p. 326, 2000.

[4] K. Coninx, R. Claes, and S. Vandael, "Anticipatory coordination of electric vehicle allocation to fast charging infrastructure," in *Advances in Practical Applications of Heterogeneous Multi-Agent Systems*Springer, Berlin, Germany, 2014.

[5] P. Morrissey, P. Weldon, and M. O'Mahony, "Future standard and fast charging infrastructure planning: an analysis of electric vehicle charging behaviour," *Energy Policy*, vol. 89, pp. 257–270, 2016.

[6] M. Yang and Zhang, "The time-of-use price of electric vehicle charging and discharging based on monte carlo algorithm," in *Proceedings of the 2018 International Conference on Big Data and Artificial Intelligence (ICBDAI 2018)*, 2018.

[7] F. Liu, "Applying fuzzy logic control to analyze real-time control for charging and discharging power of electric vehicles," *Journal of Intelligent and Fuzzy Systems*, vol. 4, pp. 1–7, 2021.

[8] F. Ni, Z. J. Liu, H. Zhang, L. Shuo, and W. Feng, "Review of research on coordinated charging of electric vehicles," *Advanced Materials Research*, vol. 860-863, pp. 1164–1172, 2014.

[9] M. T. Hussain, N. B. Sulaiman, M. S. Hussain, and J. Muhammad, "Optimal management strategies to solve issues of grid having electric vehicles (EV): a review," *The Journal of Energy Storage*, vol. 33, no. 3, Article ID 102114, 2020.

[10] S. Pourazarm, C. G. Cassandras, and A. Malikopoulos, "Optimal routing of electric vehicles in networks with charging nodes: a dynamic programming approach," in *Proceedings of the 2014 IEEE International Electric Vehicle Conference (IEVC)*, December 2019.

[11] T. Wang, C. Cassandras, and S. Pourazarm, "Energy-aware vehicle routing in networks with charging stations," *IFAC Proceedings Volumes*, vol. 473, p. 9611, 9616.

[12] S. Pourazarm and C. Cassandras, "Optimal routing of energy-aware vehicle in networks with inhomogeneous charging nodes," in *Proceedings of the 22nd IEEE Mediterranean Conference on Control and Automation*, pp. 674–679, Palermo, Italy, June 2014.

[13] M. Elgarej, M. Khalifa, and M. Youssfi, "Optimized path planning for electric vehicle routing and charging station navigation systems," *International Journal of Applied Metaheuristic Computing*, vol. 11, no. 3, pp. 58–78, 2020.

[14] S. Zhang, G. Feng, Y. Fan, S. Wan, and Y. Luo, "Large-scale electric vehicle charging path planning based on information interaction," *Qinghua Daxue Xuebao/Journal of Tsinghua University*, vol. 58, no. 3, pp. 279–285, 2018.

[15] Y. Shen, W. Fang, F. Ye, and M. Kadoch, "EV charging behavior analysis using hybrid intelligence for 5G smart grid," *Electronics*, vol. 9, no. 1, p. 80, 2020.

[16] M. Y. Yen, H. C. Chen, Y. L. Wei, and Chung, "A CMOS transmitter analog baseband for 5G mobile communication," *Electronics*, vol. 8, no. 11, p. 1319, 2019.

[17] W. Qi, B. Zla, N. Ke, C. Yifan, and W. Ming, "Dynamic resource allocation for jointing vehicle-edge deep neural network inference," *Journal of Systems Architecture*, vol. 117, Article ID 102133, 2021.

[18] T. Nishimura and H. Koizumi, "A real-time traffic flow measurement method based on inter-vehicle communication considering the volume of vehicle information," 情報処理学会研究報告高度交通システム*(ITS)*, vol. 2006, pp. 1–8, 2006.

[19] R. Sun, Y. Yang, K. W. Chiang, D. Thanh-Trung, L. Kuan-Ying, and T. Guang-Je, "Robust IMU/GPS/VO integration for vehicle navigation in GNSS degraded urban areas," *IEEE Sensors Journal*, vol. 20, no. 17, p. 1, 2020.

[20] J. Tan and L. F. Wang, "Real-time charging navigation of electric vehicles to fast charging stations: a hierarchical game approach," *IEEE Transactions on Smart Grid*, vol. 8, no. 2, pp. 846–856, 2017.

[21] G. Li, N. Cao, P. Zhu et al., "Towards smart transportation system," *Journal of Organizational and End User Computing*, vol. 33, no. 3, pp. 35–49, 2021.

[22] D. Chen, Y. Wang, and W. Gao, "Combining a gradient-based method and an evolution strategy for multi-objective reinforcement learning," *Applied Intelligence*, vol. 50, no. 10, pp. 3301–3317, 2020.

[23] M. Zaher, A. Shehab, M. Elhoseny, and F. F. Farahat, "Unsupervised model for detecting plagiarism in internet-based

handwritten Arabic documents," *Journal of Organizational and End User Computing*, vol. 32, no. 2, pp. 42–66, 2020.

[24] C. Liu, W. Jing, and C. Long, "Joint charging and routing optimization for electric vehicle navigation systems," *IFAC Proceedings Volumes*, vol. 47, no. 99, pp. 2106–2111, 2014.

[25] G. Li, Q. Sun, L. Boukhatem, J. Wu, and J. Yang, "Intelligent vehicle-to-vehicle charging navigation for mobile electric vehicles via VANET-based communication," *IEEE Access*, vol. 7, pp. 170888–170906, 2019.

[26] X. Li, Y. Xiang, L. Lyu et al., "Price incentive-based charging navigation strategy for electric vehicles," *IEEE Transactions on Industry Applications*, vol. 56, no. 5, pp. 5762–5774, 2020.