Research Article

A Day-to-Day Route Choice Model Based on Reinforcement Learning

Fangfang Wei, Shoufeng Ma, and Ning Jia

College of Management and Economic, Tianjin University, Tianjin 300072, China

Correspondence should be addressed to Ning Jia; jia_ning@tju.edu.cn

Received 10 April 2014; Accepted 28 August 2014; Published 30 September 2014

Academic Editor: X. Zhang

Copyright © 2014 Fangfang Wei 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.

Day-to-day traffic dynamics are generated by individual traveler's route choice and route adjustment behaviors, which are appropriate to be researched by using agent-based model and learning theory. In this paper, we propose a day-to-day route choice model based on reinforcement learning and multiagent simulation. Travelers' memory, learning rate, and experience cognition are taken into account. Then the model is verified and analyzed. Results show that the network flow can converge to user equilibrium (UE) if travelers can remember all the travel time they have experienced, but which is not necessarily the case under limited memory; learning rate can strengthen the flow fluctuation, but memory leads to the contrary side; moreover, high learning rate results in the cyclical oscillation during the process of flow evolution. Finally, both the scenarios of link capacity degradation and random link capacity are used to illustrate the model's applications. Analyses and applications of our model demonstrate the model is reasonable and useful for studying the day-to-day traffic dynamics.

1. Introduction

In the past decades, day-to-day traffic dynamics (DDTD) have been developed substantially in the field of transportation, which are mainly used to study the traffic fluctuations and the evolution process, rather than the final or static equilibrium state [1]. Approaches used to study DDTD are very flexible because they allow a wide range of behavior rules, levels of aggregation, and types of models to be integrated in the same modeling framework [2]. Some researchers use deterministic models or stochastic models to study the day-to-day flow evolution. Smith and Friesz et al. proposed three dynamic systems by the deterministic traffic assignment processes [3, 4]. The flow evolution trajectory can be provided explicitly in deterministic models. Daganzo and Sheffi put forward the stochastic user equilibrium to model the evolution of the traffic pattern [5]. Stochastic models mainly focus on the probability distribution of flow states [6, 7]. In some literatures, traffic dynamics are modeled as continuous time processes [3, 4]. However, it is not reasonable to adjust traffic flow continuously due to travelers' activities constraints. Therefore, more appropriate methods appear by using discrete-time systems, in which travelers repeat route

choice daily or weekly [2, 8]. The flow convergences of continuous time and discrete time are very different [9].

Another important branch of DDTD is the research about travelers' behaviors. The uncertainty inherent of travelers' behaviors increases the complexity of models about DDTD, which makes traditional models unable to represent complex DDTD well [10]. This motivates much research and efforts to solve the problem by using new methods. Among these new methods, agent-based techniques are very appropriate approaches. An agent knows its internal state, the evolution of system, the likely outcomes of its actions, and so forth, and it has some abilities, such as independent decision-making, beliefs, desires, and intentions. These characteristics ensure the scalability and robustness of agent-based models, which are important to portray the complexity of DDTD. Some scholars applied agent-based models to analyze travelers' behaviors. Rossetti modeled commuters' behaviors in the structure of beliefs, desires, and intentions (BDI) which is a classical structure in multiagent simulation. In their model, travelers could make decisions about route choice and departure time rationally [11]. Nakayama et al. proposed some models about travelers' route choice behaviors. They took into account the limitations of travelers' cognitive capabilities

and used microsimulation to examine the dynamic nature of travelers' route choice. They found that the network flow does not necessarily converge to the user equilibrium (UE) and travelers' cognitions of each route do not become homogeneous by learning [12–14]. Jha et al. assumed that each traveler used a disutility function to perceive travel time and schedule delay for evaluating the alternative travel choices, and then chose an alternative route according to the utility maximization principle [15]. Chen and Mahmassani modeled travelers' learning process by using Bayesian theory and an agent-based simulation framework to study networks' dynamic properties [16]. Klügl and Bazzan used a simple heuristic model to stimulate the process of travelers' decision-making and studied the effect of traffic forecast [17]. They set up an agent-based model to stimulate travelers' route choice and found that travelers could avoid the Braess Paradox by learning [18]. Dia applied the agent-based approach to research drivers' behaviors under the influence of real-time information [19]. Liu and Huang used multiagent stimulation to investigate the differences of route update rules in two kinds of scenarios, which are traffic information unreleased and release, respectively [20]. Illenberger et al. proposed a model to study travelers' risk-sensitive route-choice behaviors [21]. Gao and Wang explored travelers' route choice under guidance information by using microsimulation [22].

Travelers' decision-marking is an important aspect in agent-based models. DDTD is not a process of one-short choice but a process of repeated decisions. Travelers can perceive travel time and the network characteristics from their travel experience. It is hard for them to make the best choice every time due to their bounded rationality and the uncertainty of environment. They obey a stochastic route choice rule. If the travel time of a path was short in the past days, its probability to be chosen is big in the current day. On the contrary, the probability is small. These characteristics are similar with the features of the reinforcement learning (RL) theory. Under RL, if an action yielded a high payoff in the past days, the probability assigned to it increases in the current round, or the behavior associated with the action gets reinforced [23]. RL asserts that decision makers' behaviors are moved towards the direction of random choice due to the environmental uncertainty [24]. These similarities make RL very suitable to analyze travelers' behaviors. However, it has been seldom used in this field. Ben-Elia proposed a learning-based model of route choice and found that information could improve travelers' learning rate and made them more prone to risk-seeking by a laboratory experiment [25]. Wahba and Shalaby proposed a departure time and route choice model based on the Markovian decision process and RL [26–28]. Zolfpour-Arokhlo et al. formulated a route planning system by combining Q value-based dynamic programming with RL. They created a priority route plan for vehicles by studying the weights of some components in road network environments such as road safety, traffic, and weather [29]. The applications of RL in route choice still need to be further researched.

In this paper, we propose a day-to-day route choice model based on RL. Travelers' memory level, learning rate, and cognition based on their travel experience are taken into

account in the model. The decision-making and learning process of travelers follow the principles of the classical Bush-Mosteller (BM) model in RL.

The rest of the paper is organized as follows. Section 2 introduces some basic knowledge of BM model and then our day-to-day route choice model based on RL is formulated. Section 3 verifies the rationality of our model and analyzes its properties. Section 4 demonstrates the applications of our model in two traffic scenarios.

2. The Model

2.1. The Process of Travelers' Route Choice. Day-to-day route choice behavior is a repetitive decision process. Each traveler is an agent with the ability of learning and decision-making. They follow a stochastic route choice rule because of their bounded rationality and the environmental uncertainty. Travelers' route choice and travel time are not known by each other when there is no external information provided. Expected travel time (ET) and perceptive travel time (PT) are generated according to their experience. Travelers judge their choice by comparing their ET and PT after each travel and then update the probabilities that the paths will be chosen repeatedly. The characteristics of the process are similar with the features of the Bush-Mosteller (BM) model, which is a classical model in reinforcement learning (RL) and will be introduced in the next section. The process of travelers' route choice can be shown in Figure 1.

2.2. Model Formulation

2.2.1. The Bush-Mosteller Model. The Bush-Mosteller (BM) model is a classical learning model proposed by Bush and Mosteller in 1973. It consists of a learning algorithm and a stochastic decision rule. The consequences of a decision can create positive or negative stimulus, which updates the probability that the decision will be repeated [30]. Actions leading to satisfactory outcomes (i.e., outcomes that meet or exceed aspirations) will tend to be repeated in the future, whereas choices resulting in unsatisfactory experiences will be avoided. In BM model, each player focuses on his past choices and payoff but ignores all the choices and payoff of other players [31].

In the BM model, action updating takes place in two steps. Firstly, at each time t , the player i chooses an action c and calculates his stimulus by using the following formula:

$$S_c^t = \frac{u_c - A_i}{\sup |u(k) - A_i|}. \quad (1)$$

u_c is the payoff of the action c and A_i is the aspiration level of the player i at time t . $u(k)$ denotes any possible payoff if he chooses other strategies (assuming that players know all the potential payoff). The denominator in (1) represents the upper value of the set of possible differences between payoff

FIGURE 1: The process of travelers' route choice.

and aspiration. Secondly, the player i updates the selected probability of the action c as follows:

$$p_c^{t+1} = \begin{cases} p_c^t + (1 - p_c^t) l s_c^t, & \text{if } s_c^t \geq 0 \\ p_c^t + p_c^t l s_c^t, & \text{if } s_c^t < 0. \end{cases} \quad (2)$$

p_c^t is the selected probability of the action c at time t , l is the learning rate ($0 < l < 1$), and the changes of the probabilities for the actions not selected are derived from the constraint that probabilities always add up to unity.

Similarities make the BM model very suitable to research the day-to-day route choice behaviors. In the next section, a day-to-day route choice model will be proposed based on the BM model.

2.2.2. A Day-to-Day Route Choice Model. We consider a transportation network which is a fully-connected directed graph denoted as $G(N, L)$, consisting of a set of nodes N and a set of links L . Let W be the set of OD pairs, let d_w be the fixed travel demand between the OD pair $w \in W$, let R_w be the set of paths connecting the OD pair $w \in W$, let f_{kw}^t be the flow on the path $k \in R_w$ on day t , and let x_a^t be the flow on the link $a \in L$ on day t ; d , f^t , and x^t denote demand, path flow, and link flow vectors, respectively; the link-path incidence matrix is denoted as B , then $x^t = B f^t$; let C be the OD-path incidence matrix, then $d = C f^t$. The link travel time vector and the path travel time vector are denoted as T^t and G^t , respectively, then $G^t = B' T^t$ [1].

Due to bounded rationality and environmental uncertainty, travelers obey a stochastic route choice rule: the bigger the chosen probability of a path, the more likely the path is selected by the traveler. Supposing the number of optional paths for the traveler i in the OD pair $w \in W$ is N . Let P_i^t be the set of the chosen probabilities of paths for him

on day t , $P_i = \{p_1^t, p_2^t, \dots, p_n^t\}$. A random number b can be generated by computer, if

$$\sum_{j=1}^{c-1} p_j^t < b \leq \sum_{j=1}^c p_j^t, \quad (3)$$

then the traveler i chooses the path c and the flow of the path adds one.

Each traveler only knows the travel time of the path he chooses on day t . However, they can form perceptive travel time (PT) of all the paths based on their travel experience. So a traveler's PT of a path can be used as its cost during his route choice process. Due to the particularity of route choice process, we make some modifications to (1) as follows:

$$S_i^t = \begin{cases} \frac{A_i^t - M_{ic}^t}{\max \{A_i^t - M_i^t(k)\}}, & \text{if } A_i^t - M_{ic}^t \geq 0 \\ \frac{A_i^t - M_{ic}^t}{|\min \{A_i^t - M_i^t(k)\}|}, & \text{if } A_i^t - M_{ic}^t < 0. \end{cases} \quad (4)$$

S_i^t is the stimulus of the traveler i when he chooses the path c on day t . A_i^t is his expected travel time (ET) and M_{ic}^t is his perceptive travel time (PT) of the path c . $M_i^t(k)$ is his any possible PT if he chooses other path. The two denominators in (4) donate the biggest benefit and the biggest loss that the traveler may get on day t . Piecewise function can reflect the degree of the traveler's satisfaction or dissatisfaction to his route choice better.

ET and PT are formed based on travelers' travel experience. Due to the limit of human's memory, travelers remember recent travel time more clearly than older ones. So the travel time on different days has different weight on their ET

and PT. The weighted average of all the travel time that a traveler has experienced can be used to calculate his ET and PT. Let ψ be the memory level of travelers ($0 < \psi \leq 1$), then ET and PT of the traveler i can be formulated as follows:

$$A_i^t = \frac{\sum_{j=1}^{t-1} \psi^{t-1-j} T_i^j}{\sum_{j=1}^{t-1} \psi^{t-1-j}}. \quad (5)$$

A_i^t is the ET of the traveler i on day t . T_i^j is the travel time that he has experienced on day j . A_i^t is formed before the travel on day t . Consider

$$M_{ik}^t = \frac{\sum_{j=1}^t \psi^{t-j} T_i^j \xi_{ik}^j}{\sum_{j=1}^t \psi^{t-j} \xi_{ik}^j}. \quad (6)$$

M_{ik}^t is the PT of the traveler i to the path k after the travel on day t . ξ_{ik}^j is a 0-1 variable. If he chooses the path k on day j , then $\xi_{ik}^j = 1$; else $\xi_{ik}^j = 0$. Because only one path can be chosen by him on day j , then

$$\sum_{k=1}^N \xi_{ik}^j = 1. \quad (7)$$

N is the number of paths that he can choose. So the PT of the path c which is chosen by him on day t can be formulated as

$$M_{ic}^t = \sum_{k=1}^N M_{ik}^t \xi_{ik}^t. \quad (8)$$

Because the traveler i does not know the actual travel time of the paths that he does not choose on day t , it is reasonable to use PT of these paths as their cost. Because PT is the weighted average of travel time that travelers have experienced, it can reflect the stability of a path's travel time in a long time. Considering the contingency of once travel time, it is not appropriate to use the actual travel time of the path chosen on day t as its evaluation criterion, therefore we also use its PT as its cost.

Stimulus on day t can be calculated by (4) to (8). Then the probability of the path c chosen by the travel i on day t can be updated by (2). In order to ensure all probabilities always add up to unity, the probabilities of those paths not chosen by the travel i on day t are updated as follows:

$$p_k^{t+1} = \begin{cases} p_k^t - p_k^t S_i^t, & \text{if } S_i^t \geq 0 \\ \frac{p_k^t - p_k^t p_c^t S_i^t}{1 - p_c^t}, & \text{if } S_i^t < 0 \end{cases} \quad k \in N, k \neq c. \quad (9)$$

S_i^t is the stimulus of the traveler i when he chooses the path c on day t . p_k^t is the chosen probability of the path k and p_c^t is the chosen probability of the path c .

3. Model Validation and Model Properties

In the previous section, we proposed a day-to-day route choice model. In this section, some numerical stimulations

FIGURE 2: The first test network.

are made to validate the model by using three different types of test networks. The properties of the model are also analyzed.

3.1. The Simulation Program and the Three Test Networks

3.1.1. The Simulation Program. The simulation program is briefly described as follows:

Step 0. (Parameters Initialization). Generate initial chosen probability of each path for every traveler and specify the value of ψ and l .

Step 1. Generate a random number of 0 to 1 for every traveler. Travelers choose routes according to the rule mentioned in (3).

Step 2. Count the flow of each route. Compute the travel time of each route.

Step 3. Calculate every traveler's expected travel time and perceptive travel time with (5) to (8).

Step 4. Compute the stimulus of every traveler with (4).

Step 5. Update the chosen probability of each path for every traveler with (2) and (9).

Step 6. Cycle to perform Step 1 to Step 5 within the prescribed number of days.

3.1.2. The Three Test Networks. Three test networks, which are used commonly in transportation literatures, are chosen as our numerical examples. The first test network is composed of three links as shown in Figure 2. The total OD demand is 300. Links' free flow travel time and capacities are $t_1^0 = 3.5$, $t_2^0 = 3$, and $t_3^0 = 2$; $c_1 = 90$, $c_2 = 100$, and $c_3 = 70$. The average link travel time function is assumed to be the BPR function. Consider

$$T_a = t_a^0 \left[\left(1 + 0.15 \left(\frac{x_a}{c_a} \right)^2 \right) \right], \quad (10)$$

t_a^0 is the free flow travel time of the link a , c_a is the link capacity, and x_a is the link flow.

The second test network is a 3×3 grid network, which is constituted by 9 nodes, 12 links, and 6 routes [1]. Node and

FIGURE 3: The second test network.

FIGURE 4: The third test network.

TABLE 1: Link characteristics of the second test network.

Link	t_a^0	c_a	Link	t_a^0	c_a
1	2.0	360	7	1.2	180
2	1.2	360	8	1.5	240
3	1.5	240	9	1.0	150
4	1.2	180	10	3.0	360
5	1.2	360	11	1.5	240
6	1.0	150	12	1.5	240

link numbers are shown in Figure 3. The total OD demand is 9000. The free flow travel time and capacities of the 12 links are given in Table 1.

The third test network named Nguyen and Dupuis' network has 13 nodes and 19 links as shown in Figure 4. There are four O-D pairs in the network, 1-2, 1-3, 4-2, and 4-3; the numbers of paths between those O-D pairs are 8, 6, 5, and 6, respectively, as shown in Table 2. The O-D demands are deterministic and given as $q_{1-2} = 660$, $q_{1-3} = 495$, $q_{4-2} = 412$, and $q_{4-3} = 495$ [32]. The link characteristics of the network are shown in Table 3.

3.2. Model Validation. As mentioned in our model, the memory level of travelers is ψ ($0 < \psi \leq 1$). The value of ψ can affect the weight of the past travel time on travelers' current expected travel time (ET) and perceptive travel time (PT). Travelers weight recent travel time more than older ones. The smaller the ψ -value, the less the travel time of the past days

influence travelers' current ET and PT. With the increase of ψ -value, more travel time of the past days has influence on their ET and PT. When $\psi = 1$, it means that travelers can remember all travel time they have experienced and the travel time of each day has the same effect on ET and PT.

It is known that user equilibrium (UE) assumes that travelers are perfectly rational and know all information. So the conditions of reaching to UE are ideal. Obviously, $\psi = 1$ is also an ideal state in our model. If the flow distribution in our model can converge to UE under $\psi = 1$, then our model can be verified to be reasonable. The following are the stimulation results of the three test networks under the condition of $\psi = 1$, $l = 0.3$ (l is the learning rate).

In the first network, as shown in Figure 5, the flow fluctuations of the three paths become very weak on the 60th day and the network flow converges to equilibrium on the 210th day. The travel time of the three paths is approximately equal under the equilibrium. So the network flow converges to UE. Similarly, the flow of the second network converges to UE in Figure 6. The flow in the four OD pairs of the third network also converges to UE as shown in Figures 7, 8, 9, and 10, and there are almost no flow on the paths 6, 7, 8, 13, and 14 under UE.

As shown above, the three test networks can all converge to UE when $\psi = 1$ and $l = 0.3$. In order to study whether the learning rate can affect networks to converge to UE under the ideal state of $\psi = 1$, we set three different values of l : $l = 0.1$, $l = 0.5$, and $l = 0.9$. The second test network is used as the stimulation example. The corresponding day-to-day dynamics are shown in Figures 11 and 12. It can be seen that the network flows all converge to UE under different l when $\psi = 1$. And the convergence rate becomes faster with the increase of l . So the learning rate can impact the convergence rate but cannot influence networks to converge to UE when $\psi = 1$.

The results above not only illustrate that the network flow evolution based on our model can converge to UE when travelers' memory level is ideal ($\varphi = 1$), but also show that our model can demonstrate the process of flow and travel time evolution. And the network flow does not converge to UE smoothly but reach to UE after an oscillation period (as shown in Figure 6). These results validate the rationality of our model.

3.3. Model Properties. The rationality of our model has been validated in Section 3.2. In this part, its properties will be illustrated by analyzing the two parameters in the model (ψ and l). The second test network is used as the simulation example in this section.

3.3.1. Impact of Memory Level on Day-to-Day Traffic Dynamics. In order to see the impact of ψ on day-to-day traffic dynamics, we fix $l = 0.3$ and set three different values of ψ : $\psi = 0.1$, $\psi = 0.5$, and $\psi = 0.9$. The corresponding flow evolution and travel time evolution are shown in Figures 13 and 14. It can be seen clearly that the network flow does not converge to UE when $\psi = 0.1$ and $\psi = 0.5$, it converges to UE approximately as $\psi = 0.9$. And the flow fluctuations of

TABLE 2: The paths in the third test network.

OD pair	Serial number of paths	Link sequence	OD pair	Serial number of paths	Link sequence
1-2	1	1-5-7-9-11	4-2	15	3-5-7-9-11
	2	1-5-7-10-15		16	3-5-8-14-15
	3	1-5-8-14-15		17	3-5-7-10-15
	4	1-6-12-14-15		18	3-6-12-14-15
	5	2-18-11		19	4-12-14-15
	6	2-17-7-9-11	4-3	20	3-5-7-10-16
	7	2-17-7-10-15		21	3-5-8-14-16
	8	2-17-8-14-15		22	3-6-12-14-16
1-3	9	1-5-7-10-16		23	3-6-13-19
	10	1-5-8-14-16		24	4-12-14-16
	11	1-6-12-14-16		25	4-13-19
	12	1-6-13-19			
	13	2-17-7-10-16			
	14	2-17-8-14-16			

TABLE 3: Link characteristics of the third test network.

Link	t_a^0	c_a	Link	t_a^0	c_a	Link	t_a^0	c_a
1	7	300	8	13	250	15	9	300
2	9	200	9	5	250	16	8	300
3	9	200	10	9	300	17	7	200
4	12	200	11	9	500	18	14	300
5	3	350	12	10	550	19	11	200
6	9	400	13	9	200			
7	5	500	14	6	400			

$\psi = 0.1$ and $\psi = 0.5$ are strong, but that is weak relatively when $\psi = 0.9$.

The flow standard deviations (FSD) in the last 100 days can be used as the statistic to measure the flow fluctuations. We fix the value of l and increase the value of ψ from 0 to 1. Due to the length limitation of the paper, only four deterministic values of l are listed ($l = 0.3$, $l = 0.6$, $l = 0.7$, and $l = 0.9$). The corresponding changes of FSD are shown in Figure 15. After summarizing all simulation results, we find that when $l > 0.6$, the flow fluctuations weaken obviously with the increase of ψ , but when $l < 0.6$, only higher memory level ($\psi > 0.5$) can reduce the fluctuations obviously.

These phenomena above are consistent with the physical meaning of ψ : travelers can cognize travel time better with the increase of ψ . It is helpful to reduce the randomness of their route choice and make them do better choice. So a bigger ψ can weaken the flow fluctuations and promote the convergence of the network flow.

3.3.2. Impact of Learning Rate on Day-to-Day Traffic Dynamics. To see the impact of l on day-to-day traffic dynamics, we fix $\psi = 0.6$ and set three different values of l : $l = 0.1$, $l = 0.5$, and $l = 0.9$. As shown in Figures 16 and 17, we can see that the network flow converges to UE approximately only

when $\psi = 0.6$ and $l = 0.1$. The flow fluctuations are strong under $\psi = 0.6$ and $l = 0.5$. When $\psi = 0.6$ and $l = 0.9$, the flow distribution deviates from the UE for several days after the network flow converges to UE, then comes back to UE again. And the process is repeated. This phenomenon is called “cyclical oscillation” by us. The cyclical oscillation found in the flow evolution is not an accidental phenomenon, which appears in a lot of stimulations with a high learning rate (such as $\psi = 0.5$, $l = 0.9$; $\psi = 0.8$, $l = 0.9$; $\psi = 0.9$, and $l = 0.85$). The corresponding flow evolutions can be seen in Figure 18.

We fix the value of ψ and increase the value of l from 0 to 0.99 for studying the impact of l on flow fluctuations. From Figure 19, we can see that when $0 \leq \psi \leq 0.5$, the flow fluctuations strengthen with the increase of l . But when $\psi \geq 0.5$, there are two inflection points (IP) in the process of FSD changes. With the increase of l , the flow fluctuations strengthen before the first IP, and then weaken until the second IP and strengthen again after the second IP. For example, when $\psi = 0.8$, the two IP are $l = 0.5$ and $l = 0.8$, respectively. when $l \leq 0.5$ and $l \geq 0.8$, the flow fluctuations strengthen with the increase of l . But when $0.5 \leq l \leq 0.8$, the flow fluctuations weaken with the increase of l .

The impact of l on day-to-day traffic dynamics can be explained: the larger the l , the more the paths' chosen probabilities are adjusted, which strengthens the flow fluctuations

FIGURE 5: Flow evolution and travel time evolution of the first test network when $\psi = 1$, $l = 0.3$. FIGURE 6: Flow evolution and travel time evolution of the second test network when $\psi = 1$, $l = 0.3$.

FIGURE 7: Flow evolution in OD pairs 1-2 and 1-3 of the third test network when $\psi = 1$, $l = 0.3$.

FIGURE 8: Flow evolution in OD pairs 4-2 and 4-3 of the third test network when $\psi = 1$, $l = 0.3$.

FIGURE 9: Travel time evolution in OD pairs 1-2 and 1-3 of the third test network when $\psi = 1$, $l = 0.3$.

FIGURE 10: Travel time evolution in OD pairs 4-2 and 4-3 of the third test network when $\psi = 1$, $l = 0.3$.

FIGURE 11: Flow evolution in the second test network with different l -value when $\psi = 1$.

FIGURE 12: Travel time evolution in the second test network with different l -value when $\psi = 1$.

naturally. Memory level is helpful for weakening the flow fluctuations as mentioned in Section 3.3.1. So the effect of memory level and learning rate on day-to-day dynamics is opposite. Their opposite effect leads to the first IP in the process of FSD changes in Figure 19; the second IP is generated by the cyclical oscillation of flow evolution which appears under high learning rate.

4. Traffic Scenarios Analyses

In this section, in order to illustrate the applications of our model, two particular traffic scenarios are analyzed by using the third test network and some discussions are provided

based on the numerical results. The two parameters of our model are fixed as $\varphi = 0.8$, $l = 0.3$ in this section.

4.1. Link Capacity Degradation Scenario. Assuming that the capacity of the first link in the third test network reduces 50% on the 1001th day because of road maintenance, and recovers on the 2001th day [33]. The day-to-day traffic dynamics are shown in Figures 20 and 21 (different color lines represent different paths, which can be seen in Figures 7 and 8). We can see that the network flow distributions in the four OD pairs all converge to UE approximately before the link capacity reduction. After the capacity reduction, the flow distributions in all OD pairs are changed. The changes in the OD pairs 1-2 and 1-3 are bigger than those in the OD pairs 4-2 and 4-3. These are consistent with our intuition: OD pairs 1-2 and 1-3 both contain the first link and, thus, should be more severely affected by the link capacity reduction. The first link does not belong to OD pairs 4-2 and 4-3, so the two OD pairs should be influenced gently. When the first link's capacity recovers, initial network equilibrium is restored. This is because when the capacity recovers, travelers can form similar cognition on the network with that formed before the capacity reduction. So the initial network equilibrium can be restored by revoking the link capacity degradation. Other point needs to be noticed: the first link's capacity reduction strengthens the flow fluctuations in the OD pairs 1-2 and 1-3. These suggest that the capacity reduction can reduce the stability of the network.

4.2. Links' Capacities as Random Variables Scenario. Links' actual capacities may be influenced by many factors, such as bad weather, traffic accidents, pedestrians' crossing street behaviors, traffic violations, and road maintenance. All these factors can reduce links' capacities. So links' actual capacities are no longer deterministic values but random variables [34].

Let c_a be the actual capacity of the link a , let h_a be the minimum capacity, and let G_a be the maximum capacity. Assume that every link's actual capacity follows a uniform distribution, which means $c_a \sim U(h_a, G_a)$.

In order to study the impact of links' actual capacities' random variations on the day-to-day traffic dynamics, the dynamics of deterministic capacities are used as reference. Let deterministic capacities be equal to the expected values of the actual capacities, let C_a denote the deterministic capacity of the link a , and let $E(c_a)$ denote the expected value of the actual capacity of the link a . Then

$$E(c_a) = \frac{h_a + G_a}{2} = C_a. \quad (11)$$

The mean and standard deviation of paths' flow and travel time in the last 100 days are used as comparison statistics. The third test network is used as the numerical example under the condition of $\psi = 0.8$, $l = 0.3$. Deterministic capacities use links' capacities in Table 3. Take the statistical results of the OD pairs 1-3 and 4-2 as examples to show the differences between deterministic capacities and random capacities on day-to-day dynamics. The results of

FIGURE 13: Flow evolution with different ψ -value when $l = 0.3$.

deterministic capacities and random capacities are shown in Tables 4 and 5, respectively.

Because paths' mean travel time equal approximately and their standard deviation are not big in Tables 4 and 5, the network flow under the two kinds of links' capacities both converge to UE approximately. But the UE of the two cases are different, because the mean path flow distributions are not the same by comparing Tables 4 and 5. So the random variations of links' capacities change the network's paths flow equilibrium. According to the mean flow of paths in Tables 4 and 5, links' mean flow under deterministic link capacities and random link capacities can be calculated as the following formula:

$$x_a^M = \sum_p \delta_{ap}^w f_p^M. \quad (12)$$

x_a^M and f_p^M represent the mean flow of the link a and the path p in the last 100 days. If the link a belongs to the path p , then $\delta_{ap}^w = 1$, else $\delta_{ap}^w = 0$. There are 19 links in the third test network. The calculation values of links' mean flow are shown in Figure 22. It can be seen clearly that the mean flow of

links under the two kinds of links' capacities is almost equal. So the random variations of links' capacities almost do not change the equilibrium of links flow, but it can change the equilibrium of paths flow as mentioned above.

In addition, the random variations of links' capacities make the mean travel time of paths become longer, which can be explained as follows.

We assume $c_a \sim U(h_a, G_a)$ above, so the travel time of the link a is also a random variable. Let $\psi(c_a)$ be the probability density function of c_a and let T_a be the travel time of the link a . $E(T_a)$ denotes its expected travel time. Then

$$E(T_a) = \int_{h_a}^{G_a} T(c_a) \psi(c_a) dc_a = t_a^0 + 0.15 t_a^0 \frac{x_a^2}{h_a G_a}. \quad (13)$$

Because $h_a > 0$, $G_a > 0$ and $E(c_a) = (h_a + G_a)/2 = C_a$, then

$$E(T_a) > t_a^0 \left[1 + 0.15 \left(\frac{x_a}{C_a} \right)^2 \right]. \quad (14)$$

FIGURE 14: Travel time evolution with different ψ -value when $l = 0.3$.

TABLE 4: The statistical results of deterministic capacities.

OD pair	Serial number of paths	Link sequence	Mean flow	Flow standard deviation	Mean travel time	Time standard deviation
1-3	9	1-5-7-10-16	341.92	8.1509	51.975	0.5202
	10	1-5-8-14-16	34.67	3.6764	52.1231	0.352
	11	1-6-12-14-16	20.92	2.9565	52.1714	0.2675
	12	1-6-13-19	97.08	6.897	52.063	0.564
	13	2-17-7-10-16	0.24	0.5527	55.5151	0.3316
	14	2-17-8-14-16	0.17	0.4277	55.6632	0.3063
4-2	15	3-5-7-9-11	302.15	5.4926	49.7345	0.4098
	16	3-5-8-14-15	4.52	1.4318	49.9226	0.254
	17	3-5-7-10-15	37.09	3.4085	49.7746	0.4377
	18	3-6-12-14-15	1.19	1.2768	49.971	0.1818
	19	4-12-14-15	67.05	4.0982	49.7898	0.3518

FIGURE 15: The changes of FSD with the increase of ψ under deterministic values of l . FIGURE 16: Flow evolution with different l -value when $\psi = 0.6$.

FIGURE 17: Travel time evolution with different l -value when $\psi = 0.6$.

TABLE 5: The statistical results of random capacities.

OD pair	Serial number of paths	Link sequence	Mean flow	Flow standard deviation	Mean travel time	Time standard deviation
1-3	9	1-5-7-10-16	291.11	16.9657	52.5273	2.8096
	10	1-5-8-14-16	48.7	6.3794	53.3368	2.4696
	11	1-6-12-14-16	37.06	5.9286	53.5686	2.1972
	12	1-6-13-19	117.4	11.8841	53.0173	2.2925
	13	2-17-7-10-16	0.64	0.7722	55.3073	2.5127
	14	2-17-8-14-16	0.09	0.2876	56.1167	2.1885
4-2	15	3-5-7-9-11	260.96	14.4564	49.7908	2.3642
	16	3-5-8-14-15	5.55	2.1385	51.2376	1.7302
	17	3-5-7-10-15	37.03	5.1707	50.4282	1.8506
	18	3-6-12-14-15	2.45	1.6659	51.4695	1.4769
	19	4-12-14-15	106.01	14.1461	50.3646	2.9827

FIGURE 18: The cyclical oscillation in the flow evolution under some combinations of ψ and l .

The travel time of the path p can be represented as

$$T_{pw} = \sum_a \delta_{ap}^w T_a, \quad \forall p \in P_w, w \in W. \quad (15)$$

So the expected travel time of the path p is

$$E(T_{pw}) = \sum_a \delta_{ap}^w E(T_a). \quad (16)$$

Formula (14) means that when the deterministic capacities are equal to the expected values of the actual capacities, the expected travel time of links under actual capacities are longer than those under deterministic capacities. So the expected travel time of paths under actual capacities is also longer than those under deterministic capacities according to Formula (16). This is why the mean travel time of paths in Table 5 is longer than those in Table 4.

Another phenomenon is that when links' capacities change from deterministic values to random values, the flow of the paths which have the smallest mean travel time decreases by comparing Table 5 with Table 4. It is because the travel time standard deviations of these paths are bigger than

most of other paths. Then the fluctuations of the paths' travel time are stronger, respectively. In other word, the stabilities of these paths' travel time are not better. So the flow on them decreases. It suggests that travelers not only consider the mean travel time of a path, but also consider the stability of the path travel time.

5. Conclusions

In this paper, we propose a day-to-day route choice model based on reinforcement learning. Travelers' memory level and learning rate are incorporated into the model. Travelers obey a stochastic route choice rule and use their experience-based cognition to evaluate their choice and then update the chosen probabilities of paths.

The rationality of the model is verified firstly. We assume that travelers can remember all the travel time they have experienced. Under the ideal hypothesis, the flow evolution based on our model in three different types of networks can all converge to UE. It is known that UE is equilibrium under ideal condition. So our model is verified to be reasonable.

FIGURE 19: The changes of FSD with the increase of l .

FIGURE 20: Flow evolution of the third test network in the first link's capacity reduction scenario.

FIGURE 21: Travel time evolution of the third test network in the first link's capacity reduction scenario.

FIGURE 22: The mean flow of 19 links under deterministic link capacities and random link capacities.

Secondly, the properties of the model are analyzed. And we find that the network flow does not necessarily converge to UE under limited memory level. High learning rate can lead to cyclical oscillation in the process of flow evolution (as shown in Figure 18). Memory level can weaken the flow fluctuations, but learning rate can strengthen the fluctuations. When memory level is larger than 0.6, there are two inflection points in the process of flow fluctuations with the increase of learning rate (as shown in Figure 19). The first inflection point is caused by the opposite effect of memory level and learning rate on flow fluctuations. The second inflection point results from the cyclical oscillation which appears in flow evolution with high learning rate.

Two particular traffic scenarios are analyzed to illustrate the applications of our model. The first scenario is link capacity degradation. We find that the initial network equilibrium can be restored by revoking the link capacity degradation in our model. It is because when the capacity recovers, travelers can form similar cognition on the network with that formed before the capacity reduction. The second scenario is that links' capacities vary randomly because of bad

weather, accident, and so forth. The results in this scenario indicate the following: random variations of links' capacities can change the equilibrium of paths flow but cannot change the equilibrium of links flow, and these also increase the mean travel time of paths; when travelers choose route, they think about not only the mean travel time of paths, but also the stability of paths' travel time. Analyses and applications of our model demonstrate the model is reasonable and useful for studying the day-to-day traffic dynamics.

Some phenomena in the paper need to be further researched, such as the cyclical oscillation in the flow evolution as well as the influence of links' capacities' random variations on the network equilibrium. And in addition to experience, memory level, and learning rate, there are many other factors that can influence travelers' route choice, such as travel habits, travel purpose, and traffic information. The comprehensive influence of these factors on travelers' behaviors is also our future research. At the same time, empirical studies are necessary to investigate travelers' route choice behaviors.

Conflict of Interests

The authors declare that there is no conflict of interests regarding the publication of this paper.

Acknowledgments

The Project is supported by National Natural Science Foundation of China (Grant nos. 70971094, 50908155, 71101102, and 71271150).

References

- [1] X. He, X. Guo, and H. X. Liu, "A link-based day-to-day traffic assignment model," *Transportation Research B: Methodological*, vol. 44, no. 4, pp. 597–608, 2010.
- [2] D. Watling and M. L. Hazelton, "The dynamics and equilibria of day-to-day assignment models," *Networks and Spatial Economics*, vol. 3, no. 3, pp. 349–370, 2003.
- [3] M. J. Smith, "The stability of a dynamic model of traffic assignment—an application of a method of Lyapunov," *Transportation Science*, vol. 18, no. 3, pp. 245–252, 1984.
- [4] T. L. Friesz, D. Bernstein, N. J. Mehta, R. L. Tobin, and S. Ganjalizadeh, "Day-to-day dynamic network disequilibria and idealized traveler information systems," *Operations Research*, vol. 42, no. 6, pp. 1120–1136, 1994.
- [5] C. F. Daganzo and Y. Sheffi, "On stochastic models of traffic assignment," *Transportation Science*, vol. 11, no. 3, pp. 253–274, 1977.
- [6] E. Cascetta, "A stochastic process approach to the analysis of temporal dynamics in transportation networks," *Transportation Research B*, vol. 23, no. 1, pp. 1–17, 1989.
- [7] F. Yang and H. X. Liu, "A new modeling framework for travelers day-to-day route choice adjustment processes," *Transportation and Traffic Theory 2007. Papers Selected for Presentation at ISTTT17*, 2007.
- [8] K. Parry and M. L. Hazelton, "Bayesian inference for day-to-day dynamic traffic models," *Transportation Research B: Methodological*, vol. 50, pp. 104–115, 2013.
- [9] D. Watling, "Stability of the stochastic equilibrium assignment problem: a dynamical systems approach," *Transportation Research B: Methodological*, vol. 33, no. 4, pp. 281–312, 1999.
- [10] D. Watling, "Urban traffic network models and dynamic driver information systems," *Transport Reviews*, vol. 14, no. 3, pp. 219–246, 1994.
- [11] R. J. F. Rossetti, R. H. Bordini, A. L. C. Bazzan, S. Bampi, R. Liu, and D. V. Vliet, "Using BDI agents to improve driver modelling in a commuter scenario," *Transportation Research C: Emerging Technologies*, vol. 10, no. 5-6, pp. 373–398, 2002.
- [12] S. Nakayama, R. Kitamura, and S. Fujii, "Drivers' learning and network behavior dynamic analysis of the driver-network system as a complex system," *Transportation Research Record*, vol. 1676, no. 1, pp. 30–36, 1999.
- [13] S. Nakayama and R. Kitamura, "Route choice model with inductive learning," *Transportation Research Record*, vol. 1725, no. 1, pp. 63–70, 2000.
- [14] S. Nakayama, R. Kitamura, and S. Fujii, "Drivers' route choice rules and network behavior: do drivers become rational and homogeneous through learning?" *Transportation Research Record*, no. 1752, pp. 62–68, 2001.
- [15] M. Jha, S. Madanat, and S. Peeta, "Perception updating and day-to-day travel choice dynamics in traffic networks with information provision," *Transportation Research C: Emerging Technologies*, vol. 6, no. 3, pp. 189–212, 1998.
- [16] R. Chen and H. S. Mahmassani, "Travel time perception and learning mechanisms in traffic networks," *Transportation Research Record*, vol. 1894, no. 1, pp. 209–221, 2004.
- [17] F. Klügl and A. L. C. Bazzan, "Route decision behaviour in a commuting scenario: simple heuristics adaptation and effect of traffic forecast," *Journal of Artificial Societies and Social Simulation*, vol. 7, no. 1, 2004.
- [18] A. L. C. Bazzan and F. Klügl, "Learning to behave socially and avoid the braess paradox in a commuting scenario," in *Proceedings of the 1st International Workshop on Evolutionary Game Theory for Learning in MAS*, 2003.
- [19] H. Dia, "An agent-based approach to modelling driver route choice behaviour under the influence of real-time information," *Transportation Research C: Emerging Technologies*, vol. 10, no. 5-6, pp. 331–349, 2002.
- [20] T.-L. Liu and H.-J. Huang, "Multi-agent simulation on day-to-day route choice behavior," in *Proceeding of the 3rd International Conference on Natural Computation (ICNC'07)*, vol. 5, pp. 492–498, Haikou, China, August 2007.
- [21] J. Illenberger, G. Flotterod, and K. Nagel, "A model of risk-sensitive route-choice behavior and the potential benefit of route guidance," *IEEE Transactions on Intelligent Transportation Systems*, vol. 12, no. 2, pp. 384–389, 2011.
- [22] F. Gao and M.-Z. Wang, "Route choice behavior model with guidance information," *Journal of Transportation Systems Engineering and Information Technology*, vol. 10, no. 6, pp. 64–69, 2010.
- [23] R. Lahkar and R. M. Seymour, "Reinforcement learning in population games," *Games and Economic Behavior*, vol. 80, pp. 10–38, 2013.
- [24] J. G. Cross, "A stochastic learning model of economic behavior," *The Quarterly Journal of Economics*, vol. 87, no. 2, pp. 239–266, 1973.
- [25] E. Ben-Elia and Y. Shiftan, "Which road do I take? A learning-based model of route-choice behavior with real-time information," *Transportation Research A: Policy and Practice*, vol. 44, no. 4, pp. 249–264, 2010.

- [26] M. M. Wahba, *Microsimulation Learning-based Approach to Transit Assignment*, University of Toronto, 2008.
- [27] M. Wahba and A. Shalaby, "Large-scale application of MILA-TRAS: case study of the Toronto transit network," *Transportation*, vol. 38, no. 6, pp. 889–908, 2011.
- [28] M. Wahba and A. Shalaby, "Learning-based framework for transit assignment modeling under information provision," *Transportation*, vol. 41, no. 2, pp. 397–417, 2014.
- [29] M. Zolfpour-Arokhlo, A. Selamat, S. Z. Mohd Hashim, and H. Afkhami, "Modeling of route planning system based on Q value-based dynamic programming with multi-agent reinforcement learning algorithms," *Engineering Applications of Artificial Intelligence*, vol. 29, pp. 163–177, 2014.
- [30] M. W. Macy and A. Flache, "Learning dynamics in social dilemmas," *Proceedings of the National Academy of Sciences of the United States of America*, vol. 99, supplement 3, pp. 7229–7236, 2002.
- [31] S. Lasaulce and H. Tembine, *Game Theory and Learning for Wireless Networks: Fundamentals and Applications*, Academic Press, 2011.
- [32] H. Xu, Y. Lou, Y. Yin, and J. Zhou, "A prospect-based user equilibrium model with endogenous reference points and its application in congestion pricing," *Transportation Research B: Methodological*, vol. 45, no. 2, pp. 311–328, 2011.
- [33] X. Guo and H. X. Liu, "Bounded rationality and irreversible network change," *Transportation Research B: Methodological*, vol. 45, no. 10, pp. 1606–1618, 2011.
- [34] Q. Wang and W. Xu, "A user equilibrium model based on cumulative prospect theory for degradable transport network," in *Proceeding of the 4th International Joint Conference on Computational Sciences and Optimization (CSO '11)*, pp. 1078–1082, Yunnan, China, April 2011.