EFFECTS OF TRAFFIC INFORMATION ON DRIVERS' DAY-TO-DAY ROUTE CHOICES

GENARO PEQUE, JR[1], TOSHIHIKO MIYAGI[2] & FUMITAKA KURAUCHI[2]
[1]Kobe University, Japan.
[2]Gifu University, Japan.

ABSTRACT

A multi-agent route choice learning model for the microscopic simulation-based dynamic traffic assignment (DTA) is used to investigate the effects of traffic information accuracy on drivers' day-to-day route choice decisions. Using the total relative gap convergence metric to quantify the convergence speed for some chosen update cycle length intervals, the results show that a slight decrease in accuracy has a negative effect on the rate of convergence. From a learning perspective, shorter information update cycles from an advanced traveller information system induce faster convergence when compared to longer information update cycles. This implies that drivers learn faster, given the additional computational and storage costs of travel information that the system is willing to invest in. Moreover, when the update cycle length is very long, it produces a worse result compared to a scenario where drivers rely only on their own travel experiences based on the routes they have chosen.

Keywords: ATIS, day-to-day route choice, simulation-based DTA

1 INTRODUCTION

Each driver makes a decision based on his knowledge of the available alternative routes and their attributes subject to time and cognitive capacity constraints. A driver's decision depends on whether traffic information regarding available alternatives are provided. Without information, drivers' choices will be based on the knowledge they have gained from their past choices.

Traffic information is usually provided by an advanced traveller information system (ATIS) which operates through information supplied entirely within vehicles and/or traffic management centres (TMCs). ATIS is designed to assist drivers in making better route choice decisions by providing information regarding other alternative routes. Information can either be (i) descriptive such as information regarding prevailing conditions like current travel times which can be provided pre-trip, (ii) prescriptive such as suggestions of the path with the shortest travel time to a destination and (iii) feedback such as the historical records of travel times on chosen and non-chosen routes.

Experiments concerning the impact of information on drivers' route choice decisions have also been carried out by researchers. For example, repeated choice experiments were conducted by Avineri and Prasker [1] where they provided respondents with static pre-trip information regarding average expected travel times and feedback information about their chosen alternative. They showed that informed respondents preferred reliable routes compared to the non-informed respondents. Similarly, Ben-Elia et al. [2] conducted experiments where they provided respondents with dynamic en-route information describing the ranges of travel times. Their result was the opposite of the result by Avineri and Prasker as informed respondents preferred shorter and riskier routes. Concerning feedback information, Bogers et al. [3] showed that respondents who were provided with information regarding all alternative routes performed better than the respondents who used only experiential information.

ISSN: 2058-8305 (paper format), ISSN: 2058-8313 (online), http://www.witpress.com/journals
DOI: 10.2495/TDI-V2-N4-342-352

However, both benefits of the dynamic en-route and feedback information decreased as more experience was accumulated. Using all three types of information discussed above, Ben-Elia et al. [4] conducted route choice experiments to account for travel time uncertainty and information accuracy in an attempt to empirically investigate how information accuracy affects drivers' route choice decisions. Their results suggest that prescriptive information has the largest behavioural impact followed by descriptive information and then experiential feedback information.

Theoretical frameworks have also been proposed where 'rational' drivers are assumed to be maximizing their utility by choosing the best perceived route using random utility models [5, 6]. These models usually assume that each driver has knowledge of the entire network provided by the system. Another approach is to consider drivers to be individual decision-makers with 'bounded-rationality' which 'learns' the network performance by repeatedly interacting with the other drivers using the network [7]. The assumption is either that (i) prescriptive information is provided to 'informed-users' [8] or noisy feedback information is provided to 'partially informed users' (PIUs) by a TMC [9] or that (ii) experiential information is gathered by 'naïve users' (NUs) [10]. The classification of drivers into informed, PIUs and NUs by the authors was based on the interactive experiment performed by Selten et al. [11].

From a learning perspective following Peque et al. [9], we are interested in how information accuracy affects drivers' route choice decisions and how this subsequently affects the equilibrium solution such as its rate of convergence among others. Specifically, we want to investigate how drivers' route choice decisions are affected by different update cycle lengths provided by an ATIS in a simulation-based dynamic traffic assignment (DTA) and compare it with a scenario where drivers use only their day-to-day experiences. We analyse these effects using the total relative gap which is a convergence metric based on route travel times.

This article is structured as follows. In the next section, we introduce some notation, definitions and preliminary ideas presented in this article, then in Section 3, the adaptive learning algorithm and microscopic traffic simulator that will be used for the DTA is introduced. In Section 4, results and analysis of some numerical examples are presented. In the last section, we present our conclusion.

2 PRELIMINARIES

In this section, we will introduce some notation and preliminary ideas used in this article.

2.1 Drivers as individual decision-makers

Peque et al. [9] introduced the concept of PIUs and NUs where drivers are assumed to make route choice decisions independently. These users selfishly choose their routes to minimize the travel times from their origins to their destinations by interacting with the other users repeatedly as if they are playing a game. Thus, a driver may sometimes be referred to as a decision-maker, a user, a traveller or a player.

PIUs are types of drivers that can acquire travel information, such as travel times from an ATIS provided by a TMC, to make route choice decisions. The behavioural assumption for PIUs is that they solely rely on this information and are aware that this information might be noisy. This information is in the form of route travel time feedback information for all routes to their destinations from the previous day (e.g. previous iteration or previous stage). Additionally, PIUs are assumed to have bounded rationality which implies that drivers want to maximize their payoffs (e.g. choose a route with the minimum travel time to their destination), but their decisions are limited only to the information they can acquire and their memory capacity. On the other hand, NUs are types of drivers that can only acquire information of a route by actually using it. These types of users use their experiences to build assessments about available routes to their destinations and are more constrained compared to PIUs.

More formally, let us refer to a driver as a player, $i \in I$, who has a set of actions (routes), $A^i = \left\{a_1^i, \ldots, a_k^i, \ldots, a_m^i\right\}$. A player's realized payoff for choosing an action $a^i \in A^i$ is given by,

$$U^i\left(a^i\right) = u^i\left(a^i\right) + \epsilon^i\left(a^i\right), \tag{1}$$

where I is the set of players, $u^i\left(a^i\right)$ is the payoff for choosing action a^i and $\epsilon^i\left(a^i\right)$ is a random term assumed to have an unknown distribution, zero mean and bounded variance. A PIU can acquire $U^i\left(a^i\right)$ for all $a^i \in A^i$ because an ATIS provides them with these information. On the other hand, an NU can only acquire $U^i\left(a^i\right)$ for a^i if they have chosen a^i. Similar assumptions proposed above have been used before such as the models by Horowitz [12]. Horowitz proposed a stochastic model where the route travel times have dependent and independent random terms. In the same paper, he also proposed a model where drivers can only acquire information about the routes that they have actually used. A closely related model is the traditional stochastic user equilibrium (SUE) model [13, 14] which has a similar form as eqn (1). However, in the traditional SUE model, all players have the same estimated payoff values for each action which implies that all players have the same route choice probabilities. Contrarily, our model assumes that players form individual action value estimates for each of their actions. Additionally, the traditional SUE model assumes that the probability distribution of the random term is known.

Now, consider a discrete time process of vectors $\left\{U_t\right\}_{t>0}$. At stage t, each player, having observed the past realizations $\left\{U_1, \ldots, U_{t-1}\right\}$, chooses an action $k \in A^i$. Each player's objective is then to maximize the expected payoff defined by,

$$\mathbb{E}\left[\liminf_{t \to +\infty} \frac{1}{t} \sum_{s=1}^{t-1} U_s^i\right], \tag{2}$$

by selecting an action repeatedly. The probability that a player i chooses an action k at time t is represented by the mixed strategy, $\pi_t^i(k) = \Pr\left[a_t^i = k\right] \in \Sigma^i$. A (mixed) Nash equilibrium is achieved when each player plays a best response to the opponents' strategies so that,

$$u^i\left(\pi_*^i, \pi_*^{-i}\right) \geq u^i\left(\pi^i, \pi_*^{-i}\right), \forall \pi^i \in \Sigma^i. \tag{3}$$

An $\epsilon -$ Nash equilibrium of a game is then defined as an action profile that satisfies the condition,

$$u^i\left(\hat{\pi}_*^i, \pi_*^{-i}\right) \geq u^i\left(\pi^i, \pi_*^{-i}\right) - \epsilon, \epsilon > 0, \forall \pi^i \in \Sigma^i. \tag{4}$$

In transportation, the payoff is usually assumed to be related to the route travel times. Since routes are made up of links, route travel times can be decomposed into link travel times which are functions of link flows. More formally, let us define the flow conservation equations. For simplicity, we restrict our attention to a single-origin destination (OD) transportation network connected by several routes. The action set of each player corresponds to the set of available routes in their OD. Path (route) flows are denoted by an m-dimensional vector $h = (h_1, \ldots, h_k, \ldots, h_m)$. Let L be a set of links, $\{f_l\}$ be the flow on the link $l \in L$ and $\{\delta_{l,k}\}$ an element of the link–path incidence matrix. A visit to a path, k, at time t by a player is described by the indicator function, $\mathbb{I}\{\cdot\}$, whose value is 1 if the statement in the parenthesis is true and 0 otherwise. Then, the cumulative and relative frequencies of visit up to time t are given by,

$$Z^i_{k,t} = \sum_{\tau}^{t} \mathbb{I}\left\{a^i_t = k\right\} \text{ and } z^i_{k,t} = \frac{1}{t}\sum_{\tau}^{t} \mathbb{I}\left\{a^i_t = k\right\}, \tag{5}$$

and the aggregated path and link flows at time t are defined as follows:

$$\sum_{i \in I} \mathbb{I}\left\{a^i_t = k\right\} = h_{k,t}, \forall k \in A \text{ , and } \sum_{k \in A} \delta_{l,k} h_{k,t} = f_{l,t}, \forall l \in L. \tag{6}$$

In our approach, a link delay function is not necessary. However, if the realized link travel time at time t, $C_{l,t}$, is necessary, then the payoff for path k would be given by

$$U^i_{k,t} = -\sum_{l \in L} \delta_{l,k} \left(\omega^i C_{l,t} + F_l\right), \tag{7}$$

where ω^i and F_l denote the value of time for player i and the fare imposed on link $l \in L$, repectively.

2.2 Travel information

Travel information sent by the TMC to the drivers in the network through the ATIS is assumed to be collected by traffic detectors in each link at the end of each update cycle in each iteration. The TMC processes current information and sends it to the ATIS as feedback information for the driver. The feedback information received by each driver are averaged route travel times from his/her origin to his/her destination which depends on his/her departure time and the duration of his/her travel in the previous day. This is shown in Fig. 1.

The description and Fig. 1 is only applicable to the PIUs. For the NUs, it is much simpler. NUs only use the exact travel time of the route they have used and have no information of the

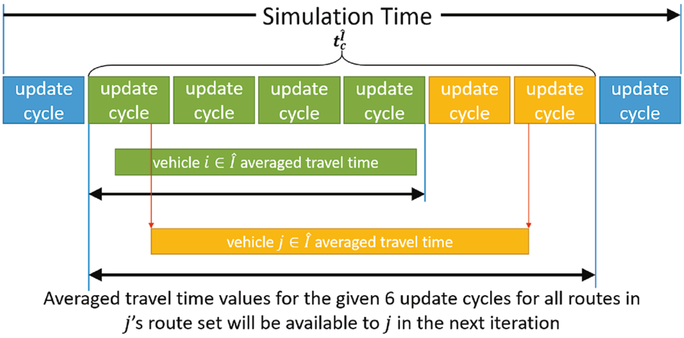
Simulation Time
t_c^1
update cycle
update cycle
update cycle
update cycle
update cycle
update cycle
update cycle
update cycle
vehicle $i \in \hat{I}$ averaged travel time
vehicle $j \in \hat{I}$ averaged travel time
Averaged travel time values for the given 6 update cycles for all routes in
j's route set will be available to j in the next iteration

Figure 1: PIUs averaged travel times based on their update cycle length (author's own construction).

travel times for their alternative routes. Both the PIUs and NUs then use an adaptive learning process, which will be described in Section 3, to estimate their route travel times and update their route choice probabilities.

2.3 Convergence metric

Although the stopping criterion that will be used during the traffic simulation is the number of iterations, a convergence metric is used to measure the effects of different update cycle lengths on the rate of convergence. The metric is called the total relative gap [15] which quantifies how close the solution is to equilibrium. Its measure reflects the summation of the differences between the average route travel times and minimum route travel times. More formally, the total relative gap is given by,

$$gap_{rel} = 1 - \left[\sum_{t_\psi} \sum_{\omega \in \Omega} F_\omega^{t_\psi} \min_{k \in K_\omega} c^{t_\psi}(k) \Big/ \sum_{t_\psi} \sum_{\omega \in \Omega} \Big(\sum_{\hat{k} \in K} f^{t_\psi} c^{t_\psi}(\hat{k}) \Big) \right], \tag{8}$$

where t_ψ is the ψth update cycle at stage t, ω is the OD pair in the in the OD pair set Ω, k is a route in the route set $K_{\omega\in\Omega}$, $f^{t_\psi}(k)$ is the flow on route k at t_ψ, $F_\omega^{t_\psi}$ is the total flow in OD pair ω at t_ψ and $c^{t_\psi}(k)$ is the experienced route travel time on route k at t_ψ. The intuition behind the total relative gap is that if all used routes have travel times very close to the shortest route travel time, the total relative gap will be close to zero. In most DTA applications, the solution is assumed to have converged to an equilibrium solution when the total relative gap is less than a pre-specified tolerance level. Since the total relative gap measures experienced route travel times, this is only applicable to the NU case. For each PIU case (e.g. different update cycle lengths), an approximate total relative gap is used to quantify this effect on drivers' route choice decisions. The approximate total relative gap averages the route travel times of each driver in the range of update cycles it belongs to while the driver was in the network as shown in Fig. 1. Thus, varying accuracy of route travel times will be acquired based on the length of the update cycles.

3 THE LEARNING ALGORITHM

3.1 The basic elements of the algorithm

In order to solve eqn (2), each player adaptively learns in two steps. The first step is the payoff estimation called the payoff learning where he/she updates his/her route travel time estimates based on his/her experience or based on the online information provided by the TMC through the ATIS. At each stage, a player creates an assessment of the payoff performance of their actions, which necessitates the use of the stochastic approximation theory to conduct the assessment.

The payoff dynamics for NUs are generally defined by the following updating equations:

$$Q_{k,t}^i = Q_{k,t-1}^i + \lambda_t \mathbb{I}\{a_t^i = k\}\left(U_t^i - Q_{k,t-1}^i\right), k \in A^i. \tag{9a}$$

For the PIUs, it reduces to a simpler form

$$Q_{k,t}^i = Q_{k,t-1}^i + \lambda_t \left(U_t^i - Q_{k,t-1}^i\right), k \in A^i, \tag{9b}$$

where $\{\lambda_t\}_{t>0}$ is a deterministic system satisfying the conditions

$$\sum_t \lambda_t = +\infty \text{ and } \sum_t (\lambda)^2 < +\infty . \tag{9c}$$

Normally, $\lambda_t = \left(\eta^i + Z^i_{k,t}\right)^{-\rho}, \eta^i > 0, \rho \in (0.5,1]$, so that the effect of each successive period diminishes and the effects of the random terms eventually vanish as $t \to \infty$. In general, $Q^i_{k,t} \to \mathbb{E}\left[U^i_t \mid a^i_t = k\right]$ with probability 1 if each player uses an ϵ – greedy action selection strategy [16]. A player's estimated expected payoff for selecting the action $a^i \in A^i$ is defined as

$$\bar{r}^i_t\left(a^i, Q^i_t\right) = \bar{r}^i_{t-1}\left(a^i, Q^i_t\right) + \frac{1}{t}\left(Q^i_t\left(a^i\right) - \bar{r}^i_{t-1}\left(a^i, Q^i_t\right)\right). \tag{10}$$

Then, the average regret of a player defined as $\bar{R}^i_t\left(a^i\right) = \bar{U}^i_t - \bar{r}^i_t\left(a^i, Q^i_t\right)$ where $\bar{U}^i_t = \frac{1}{t}\sum_t U^i_\tau$ eventually vanishes as $t \to \infty$.

The second step is the updating of route choice probabilities based on the route travel time estimates given by eqns (9a–9c) and (10) in the first step. It is the so-called strategy learning and is redundant for a stationary process, though it is inevitable for a non-stationary process. The combined payoff learning and strategy learning utilized in this article is called an actor-critic process which belongs to a general class of adaptive processes called the generalized weakened fictitious play (GWFP) [17]. Since there is no assumption on the distribution of the random term, there is no explicit relation between a player's payoff and mixed strategy. Hence, a model-free [18] or an ϵ-greedy approach [19] is effective in this case. An ϵ-greedy approach is where the action with the highest estimated payoff is selected with probability $(1-\epsilon_t)$ while the other actions are selected with probability ϵ_t where $1 > \epsilon_t > 0$. An important feature in these approaches is for each action to be selected infinitely often to maintain accurate payoff estimates. Chapman et al. [19] applied an ϵ-greedy action selection to Q-learning fictitious play which belongs to the class of GWFP. We slightly use a different approach by defining players' action probabilities as

$$\pi^i_t\left(a^i\right) = \left(1-\alpha_t\right)\pi^i_{t-1}\left(a^i\right) + \alpha_t \beta^i_t\left(a^i, Q^i_t\right), \forall a^i \in A^i, \tag{11a}$$

where $\alpha_t = \frac{1}{t}$,

$$\beta^i_t\left(a^i, Q^i_t\right) = exp\left\{Q^i_t\left(a^i\right) / \mu^i_t\right\} / \sum_{\tilde{a}^i \in A^i} exp\left\{Q^i_t\left(\tilde{a}^i\right) / \mu^i_t\right\} \text{ and } \mu^i_t = \left|\bar{R}^i_t\left(a^i\right)\right|. \tag{11b}$$

Since $\mu^i_t \to 0$ as $t \to \infty$, the choice probability of selecting the best action of each player is expected to asymptotically approach 1. This implies that all actions always have a positive probability of being chosen and thus will be selected infinitely often to maintain accurate payoff estimates.

3.2 The multi-agent Q-learning algorithm

We now integrate the entire process mentioned in Section 3.1 into a single process.

1. Initialization: At time $t = 0$, each player randomly selects the action, a^i_t.
2. Perform the traffic simulation: Each player receives the payoff U^i_t based on:

   i. For NUs: Their actual experienced travel time.
   ii. For PIUs: The information given by the ATIS.
3. Payoff learning: Payoff estimation is performed according to eqn (9b for PIUs and 9a for NUs).
4. Strategy learning: Strategy updating is performed according to eqn (11a–11b).
5. Action selection: Each player selects an action based on π_t^i.
6. Stopping criterion: Repeat steps 2–5 until a stopping criterion is met.

3.3 The dynamic traffic assignment

The DTA is conducted using the simulation of urban mobility (SUMO) microscopic traffic simulator and the adaptive learning, including route choice, is performed using eqns (9a–9c) and (11a–11b) as shown in Fig. 2.

SUMO uses a continuous version of the cellular automaton model [20] which was first proposed for single lanes by Ref. [21]. A cellular automaton model, commonly known as the Nasch model, is an extremely simplified program for the simulation of complex transportation systems where the road is subdivided into discrete cells of similar sizes. Each cell is empty or occupied by one vehicle, with discrete speed v varying from zero to v_{max}, where v_{max} is the maximum (or desired) speed of the vehicle. A cell only exchanges transported units with its neighbouring cells directly within one time step. In a Nasch model, the motion of each vehicle, i, is described by the following rules:

1. Acceleration: if $v^i < v_{max}^i$, then $v^i = v^i + 1$
2. Deceleration: if $v^i > gap_v^i$, then $v^i = gap_v^i$, where gap_v^i is the vehicle gap
3. Randomization: if $v^i > 0$, then $v^i = v^i - 1$ with braking probability p_b^i
4. Movement: $x = x + v^i$, where x is the length of each cell

The braking probability represents the rate of speed reduction of a vehicle even when there is no vehicle in front of it. Krauß [20] extended this by using a notion of safe velocity to restrict vehicles from sudden stops which occurs in the Nasch model. Although Krauß's model is more general and realistic, it involves floating point arithmetic and division which makes it slower than the Nasch model.

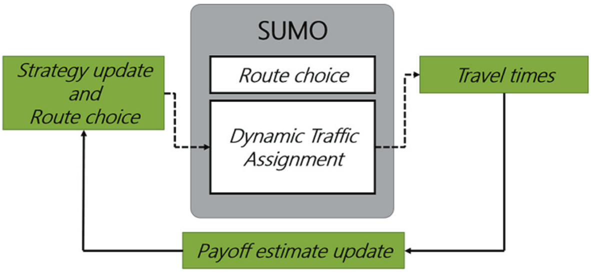

Figure 2: Simulation-based DTA with SUMO (author's own construction).

4 NUMERICAL SIMULATION

In this section, the settings and results from the numerical simulations will be shown and discussed.

4.1 Numerical simulation settings

The numerical simulation was carried out using a modified Sioux Falls network shown in Fig. 3.

The modified Sioux Falls network is made up of 14 nodes, 42 links and 48 OD pairs. There are 34,000 travellers distributed uniformly at random on all OD pairs with fixed departure times. The PIU cases were carried out for 100 iterations, each using update cycle lengths 300, 500, and 5,000 simulation seconds where the 300 update cycle length is used as the baseline for comparison. Similarly, 100 iterations were used for the NU case. Moreover, a value of η =1 and ρ =1 was used for both the PIU and NU cases.

4.2 Numerical simulation results

In the simulation, we measured the mean link and route travel times. Figure 4 shows the link and route information for OD pair 15 which is composed of the OD nodes 2 and 13, respectively. It shows that although the route travel times differ in the first 20 iterations, it suddenly dropped to a level where it became stable. It can be noticed that the route count is actually steadily increasing which implies that it was the effect of the other ODs rather than the effect of drivers changing routes abruptly within OD 15. Additionally, it also shows that the vehicles did choose the route with the lowest travel time which is route 47.

From the perspective of the drivers, Fig. 5 shows that the mean route probability is asymptotically converging to 1. Additionally, the effect of the other ODs on the average payoff, $\bar{U}$, of the drivers can be observed. It shows that even when most drivers in OD 15 have correctly chosen route 47 which has the lowest travel time, the route changes by the other drivers in the other ODs affected the travel times of some links in this route.

Figure 6 shows how the update cycle length affects the convergence of the drivers' route choices. The 300 update cycle length shows a faster rate of convergence than the rest of the PIU cases including the NU case. Moreover, a slight increase in the update cycle length (i.e.

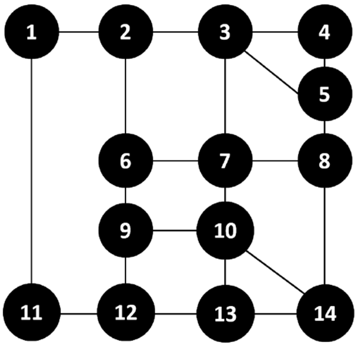

Figure 3: A modified Sioux Falls network (author's own construction).

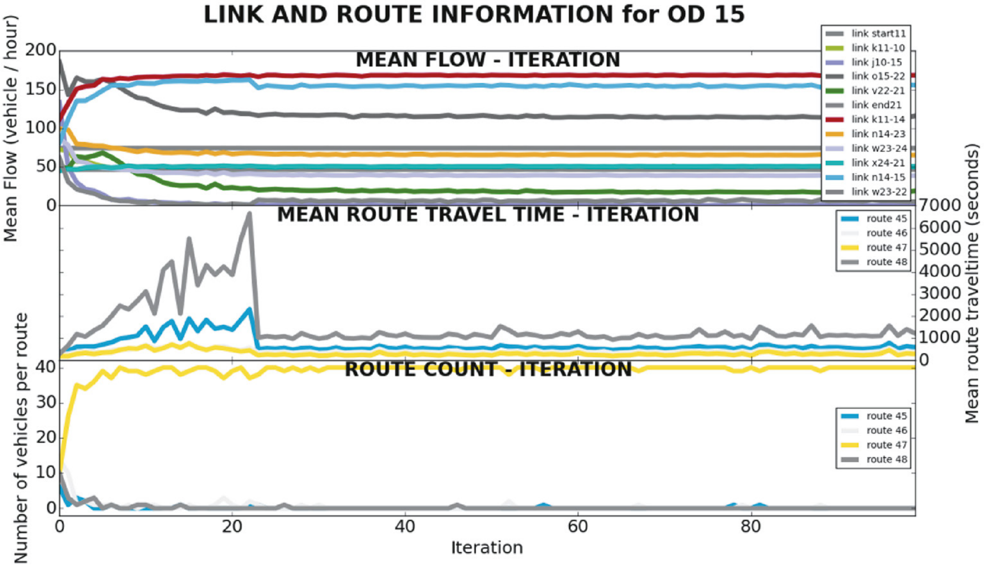

Figure 4: Link and route information for OD 15 of a PIU case with 300 update cycle length (author's own construction).

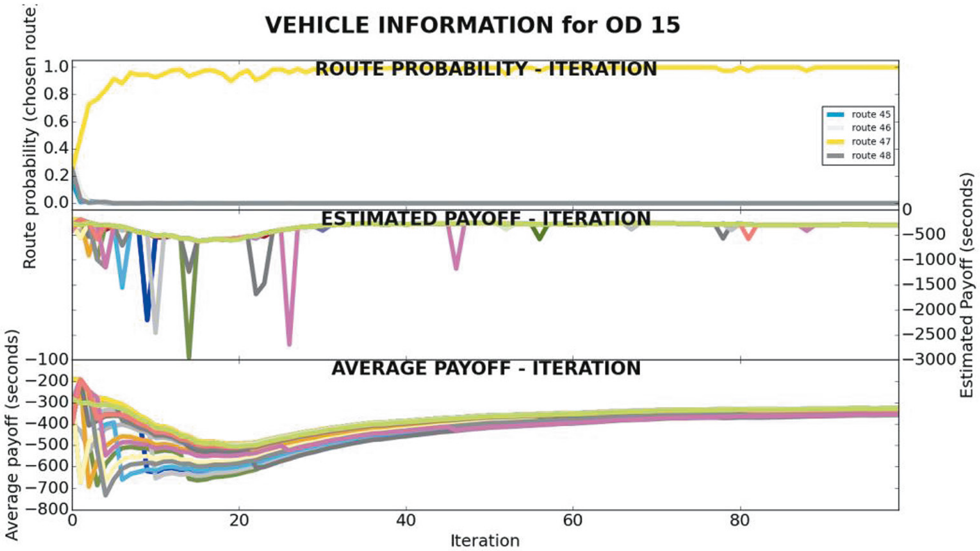

Figure 5: Vehicle information for OD 15 of a PIU case with 300 update cycle length (author's own construction).

from 300 to 500 simulation seconds) negatively affected the rate of convergence of the DTA. Interestingly, the PIU case with a 5,000 update cycle performed worse than the NU case. This implies that drivers are better off using their own experiences to make route choices when the traffic information provided to them is highly inaccurate. However, as experience is gained by both the PIUs and the NUs, the differences between them decreases.

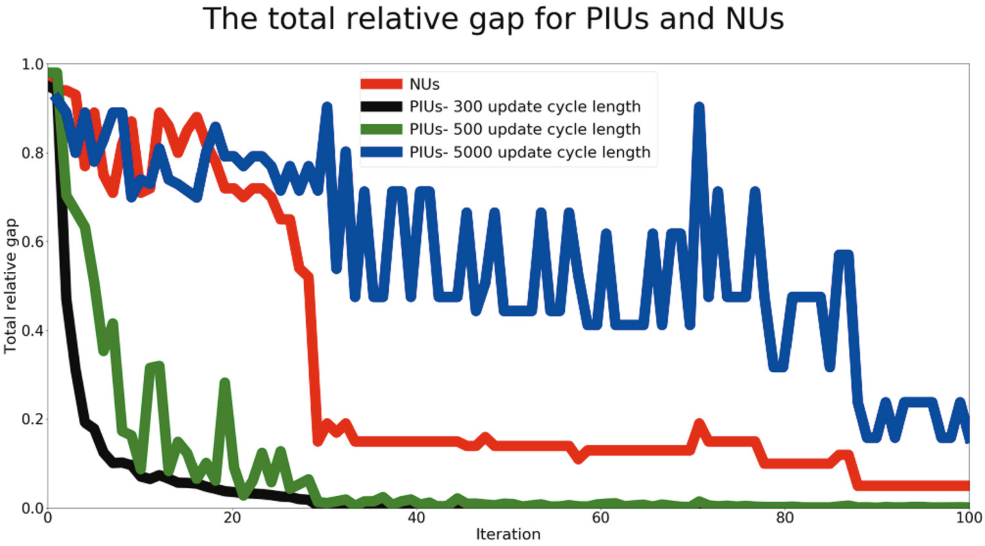

Figure 6: The total relative gap for the PIUs and the NUs (author's own construction).

5 CONCLUSION

Using a microscopic traffic simulator and an adaptive learning algorithm, we have shown the effects of traffic information on drivers' day-to-day route choice decisions. Simulation results show that the rate of convergence of the DTA is affected by the information that is given to drivers in the network. Additionally, it shows that a slight decrease in the accuracy of traffic information has a negative effect on the rate of convergence. Furthermore, simulation results show that it is better for drivers in the network to rely on their experiences rather than follow a highly inaccurate traffic information. However, our results also show that as more experience is gained by both the PIUs and NUs, the benefits of traffic information decreases since in all cases the total relative gap is decreasing as the number of iterations increased.

6 ACKNOWLEDGEMENT

This work was supported by JSPS KAKENHI Grant Number JP17H03323.

REFERENCES

[1] Avineri, E. & Prashker, J.N., The impact of travel time information on travelers' learning under uncertainty. *Transportation,* **33**, pp. 393–408, 2006. DOI: 10.1016/0191-2615(92)90017-Q.

[2] Ben-Elia, E., Erev, I. & Shiftan, Y., The combined effect of information and experience on drivers route-choice behavior. *Transportation*, **35**, pp. 165–277, 2008. DOI: 10.1037/h0092987.

[3] Bogers, E.A.I. & van Zuylen, H.J., The influence of reliability on route preferences in freight transport. Paper Presented to the 84th Annual Meeting of the Transportation Research Board, Washington and published in Transportation Research Record, 2005.

[4] Ben-Elia, E., Di Pace, R., Bifulco, G.N. & Shiftan, Y., The impact of travel information's accuracy on route-choice. *Transportation Research Part C*, **26**, pp. 146–159, 2013. DOI: 10.1016/j.trc.2012.07.001.

[5] Cascetta, E., *Transportation Systems Engineering: Theory and, Methods,* Kluwer Academic Publishers: Dordrecht, 2001.

[6] Prashker, J. & Bekhor, S., Route-choice models used in the stochastic user equilibrium problem: a review. *Transport Reviews*, **24**, 437–463, 2004. DOI: 10.1016/S0191-2615(00)00045-X.

[7] Miyagi, T., Multiagent learning models for route choices in transportation networks: an integrated approach of regret-based strategy and reinforcement learning. *Proceedings of the 11th International Conference on Travel Behavior Research*, Kyoto, 2006.

[8] Miyagi, T. & Peque, G., Informed user algorithm that converge to a pure Nash equilibrium in traffic games. *Procedia – Social and Behavioral Sciences*, **54**, pp. 438–449, 2012. DOI: 10.1016/j.sbspro.2012.09.762.

[9] Peque, G., Miyagi, T. & Kurauchi, F., Adaptive learning algorithms for simulation-based dynamic traffic user equilibrium. *International Journal of Intelligent Transportation Systems Research*, **16**(**3**), pp. 215–226, 2018. DOI: 10.1023/A:1007678930559.

[10] Miyagi, T., Peque, G. & Fukumoto, J., Adaptive learning algorithms for traffic games with naive users. *Procedia – Social and Behavioral Sciences*, **80**, pp. 806–817, 2013. DOI: 10.1016/j.sbspro.2013.05.043.

[11] Selten, R., Schreckenberg, M., Chmura, T., Pitz, T., Kube, S., Hafstein, S., Chrobok, R., Pottmeier, A. & Wahle, J., Experimental investigation of day-to-day route-choice behaviour and network simulations of autobahn traffic in North Rhine-Westphalia. *Human Behaviour and Traffic Networks*. eds. A. Schreckenberg & R. Selten, Springer: Berlin Heidelberg, pp. 1–21, 2004.

[12] Horowitz, J.L., The stability of stochastic equilibrium in a two-link transportation network. *Transportation Research Part B: Methodological*, **18**, pp. 13–28, 1984. DOI: 10.1016/0191-2615(84)90003-1.

[13] Daganzo, C. & Sheffi Y., On stochastic models of traffic assignment. *Transportation Science*, **11**, 253–274, 1977. DOI: 10.1287/trsc.11.3.253.

[14] Sheffi, Y. & Powell, W., An algorithm for the equilibrium assignment problem with random link times. *Networks*, **12**, 191–207, 1982. DOI: 10.1016/0191-2615(81)90046-1.

[15] *Dynamic Traffic Assignment: A Primer*, Transportation Network Modeling Committee, pp. 11, 2011.

[16] Singh, S., Jaakkola, T., Szepesvari, C. & Littman, M., Convergence results for single-step on-policy reinforcement-learning algorithms. *Machine Learning* **38**(**3**), pp. 287–308, 2000. DOI: 10.1023/A:1007678930559.

[17] Leslie, D. & Collins, E., Generalised weakened fictitious play. *Games and Economic Behaviour,* **56**, pp. 285–298, 2006. DOI: 10.1016/j.geb.2005.08.005.

[18] Marden, J., Young, P., Arslan, G. & Shamma, J., Payoff-based dynamics for multiplayer weakly acyclic games. *SIAM Journal on Control and Optimization*, **48**(**1**), 2009. DOI: 10.1137/070680199.

[19] Chapman, A., Leslie, D., Rogers, A. & Jennings, N., Convergent learning algorithms for unknown reward games. *SIAM Journal on Control and Optimization* **51**(**4**), pp. 3154–3180, 2013. DOI: 10.1137/120893501.

[20] Krauß, S., Towards a unified view of microscopic traffic flow theories. *Proceedings of the 8th IFAC/IFIP/IFORS Symposium*, Chania, Greece, 16–18 June, eds. M. Papageorgiou & A. Pouliezos, Vol. 2, Elsevier Science, 1997.

[21] Nagel, K. & Schreckenberg, M., A cellular automaton model for traffic flow. *Journal de Physique I*, **2**, p. 2221, 1992. DOI: 10.1051/jp1:1992277.