TRANSPORT
ISSN 1648-4142 / eISSN 1648-3480
Special Issue on Travel Demand Management / 2014 Volume 29(3): 296–306
doi:10.3846/16484142.2014.953997

EVALUATING STAGGERED WORKING HOURS USING A MULTI-AGENT-BASED Q-LEARNING MODEL

Min Yang, Dounan Tang, Haoyang Ding, Wei Wang, Tianming Luo, Sida Luo

School of Transportation, Southeast University, China

Submitted 14 October 2013; resubmitted 11 March 2014; accepted 6 April 2014

Abstract. Staggered working hours has the potential to alleviate excessive demands on urban transport networks during the morning and afternoon peak hours and influence the travel behavior of individuals by affecting their activity schedules and reducing their commuting times. This study proposes a multi-agent-based Q-learning algorithm for evaluating the influence of staggered work hours by simulating travelers' time and location choices in their activity patterns. Interactions among multiple travelers were also considered. Various types of agents were identified based on real activity–travel data for a mid-sized city in China. Reward functions based on time and location information were constructed using Origin–Destination (OD) survey data to simulate individuals' temporal and spatial choices simultaneously. Interactions among individuals were then described by introducing a road impedance function to formulate a dynamic environment in which one traveler's decisions influence the decisions of other travelers. Lastly, by applying the Q-learning algorithm, individuals' activity–travel patterns under staggered working hours were simulated. Based on the simulation results, the effects of staggered working hours were evaluated on both a macroscopic level, at which the space–time distribution of the traffic volume in the network was determined, and a microscopic level, at which the timing of individuals' leisure activities and their daily household commuting costs were determined. Based on the simulation results and experimental tests, an optimal scheme for staggering working hours was developed.

Keywords: Q-learning; multi-agent; space–time distribution; activity–travel behavior; staggered working hours.

Reference to this paper should be made as follows: Yang, M.; Tang, D.; Ding, H.; Wang, W.; Luo, T.; Luo, S. 2014. Evaluating staggered working hours using a multi-agent-based Q-learning model, *Transport* 29(3): 296–306. http://dx.doi.org/10.3846/16484142.2014.953997

Introduction

Beginning in the early 1970s, the concept of Travel Demand Management (TDM) was introduced in Europe and the US to describe strategies and policies for reducing travel demand or redistribute it in space or time. There is a broad range of TDM measures, including pricing tolls with respect to peak hours, improving public transportation, encouraging carpooling, staggering work hours and others. Staggered working hours is proposed to mitigate congestion through adjusting workers' starting and quitting times to have differing work schedules which can flatten peak congestion hence lowering individuals' commuting time. Staggered working hours has the potential to mitigate congestion and to alleviate the excessive demands made on the transport infrastructure (D'Este 1985). However, whether a firm choose to adopt this policy or not is concerned with the trade-off between productivity and congestion (Mun, Yonekawa 2006). Recent research has shown that staggered working hours might be welfare enhancing (Gutiérrez-i-Puigarnau, Van Ommeren 2012). Transport policy makers typically justify these measures by arguing that they will provide user benefits and alleviate traffic jams. The Activity-Based Modeling (ABM) framework was originally developed in response to a call for more realistic travel demand models capable of analyzing a wider range of transportation policies. These models specify the daily pattern of activity and travel at a disaggregate level for modeled regions and generally have a more behavioral basis than aggregate travel demand models. Through researchers' persistent efforts, comprehensive, operational activity-based travel demand models have become available (Timmermans *et al.* 2002). The academic research community has recently started to address a new challenge: how to develop practical activity-based travel demand models and evaluate the impacts of TDM policies using these models.

Corresponding author: Haoyang Ding
E-mail: *dinghy123@163.com*

Copyright © 2014 Vilnius Gediminas Technical University (VGTU) Press
http://www.tandfonline.com/TRAN

Taylor & Francis
Taylor & Francis Group

The traditional models used to investigate activity–travel patterns can be classified into three types: econometric models, Computational Process Models (CPM) and hybrid models. Econometric models link individual or household sociodemographics, transportation policies and other environmental factors to their activity–travel patterns. Econometric models, including discrete choice models such as multinomial logit models and nested logit models, have proven to be powerful tools for activity–travel analysis (Ettema *et al.* 2007; Yang 2007). Computational process models focus on using context-dependent choice heuristics to model an individual's decision process. The techniques used in more recent studies have included decision trees, neural networks and Bayesian networks (Benenson *et al.* 2008; Bhat *et al.* 2004). Both econometric models and CPMs have their drawbacks. Econometric models are limited in their ability to simulate travel–activity scheduling behavior, while CPMs lack a basis in statistical error theory, which makes it difficult to generalize the outcomes and apply them to policy evaluation (Dia 2002). These limitations have led to the formulation of hybrid models that integrate econometric models and CPMs. For example, Charypar and Nagel (2005) combined a decision tree with parametric modeling, and Janssens *et al.* (2007) incorporated random utility maximization into an activity scheduling model.

In addition to the above three types of models, new approaches have been introduced into the field. Among these new approaches is the agent-based model. An agent can be thought of as computer surrogate for a person or a process that fulfills a stated action. The flexibility and computational advantages of agent-based models have made them powerful tools in modeling complex systems, such as transportation systems. Holmgren *et al.* (2012) presented the Transportation And Production Agent-based Simulator (TAPAS), which is an agent-based model for simulation of trip chains. Benenson *et al.* (2008) presented PARKAGENT, a spatially explicit agent-based model for parking in the city. One branch of this research area is agent-based reinforcement learning in activity–travel choice simulation. An example is ALBATROSS, which is short for A Learning-BAsed Transportation-oriented Simulation System (Arentze, Timmermans 2004). ALBATROSS is a rule-based multi-agent system that predicts activity patterns. Researchers have proposed several activity scheduling systems for modeling space–time constraints and choice behavior within such constraints and for modeling the adaptive behavior of individuals in response to transportation control measures (Adler *et al.* 2005; Arentze, Timmermans 2008; Janssens *et al.* 2007; Hunt *et al.* 2012). In these systems, the theory of reinforcement learning is often applied to describe individual choice processes in complex environments.

To evaluate a TDM policy, aggregated traffic forecasting models combined with network operation methods have long been applied to assessing the effectiveness of TDM in alleviating congestion (Hug *et al.* 1997). However, TDM is capable of more than simply reducing congestion. Recently, researchers have begun to evaluate the effectiveness of TDM polices in more comprehensive ways. For instance, Creutzig and He (2009) investigated the impacts of TDM on air pollution, noise, climate change and traffic accidents and showed that a road charge implemented under TDM could not only address congestion but also benefit the environment. Researchers have also examined the individual benefits of TDM. The so-called logsum approach, which is rooted in random utility-based discrete choice theory (Ben-Akiva, Lerman 1985), has served for over two decades as the dominant means to assess user benefits, conceptualized as differences in expected consumer surplus (Dong *et al.* 2006). However, researchers have found that the econometric assumptions underlying the most basic logsum formulations may often be too strict. Thus, they have proposed ways to relax those assumptions while maintaining closed-form solutions or at least tractable formulations that can be solved by means of simulation (Cherchi, Polak 2005). Chorus and Timmermans (2009) attempted to relax some of the behavioral assumptions under the main framework of the logsum approach and focused on the assumed level of travelers' awareness of changes occurring. Several types of modeling methods and ideas have been proposed for the application of these models to the evaluation of certain TDM policies (Bellemans *et al.* 2012), but there has been a lack of practical applications.

As mentioned above, activity-based models specify the daily pattern of activity and travel at a disaggregate level. However, the following problems in evaluating TDM policies using activity-based models require further research:

- Traffic space–time distribution features are extremely useful in TDM evaluation because individuals' space and time decisions always influence each other and change simultaneously. However, many previous traffic forecasting models were designed to simulate individuals' space and time decisions separately. This could potentially make the models insensitive to changes in the activity scheduling process, such as when a TDM policy is introduced.
- Many activity-based models are better at replicating observed outcomes than at describing how those outcomes were reached because the interactions among individuals both before and during trips are often neglected. Traffic space–time distribution features are actually the results of interactions.
- Recent studies on agent-based simulation have mainly focused on the modeling approach and testing model sensitivity using presumptive data; the practical application of the model is seldom studied. In addition, with respect to evaluating TDM, traditional aggregated traffic forecasting models mainly focus on an entire transportation system, while the logsum approach focuses on individual interests.

Therefore, this research was conducted to develop a multi-agent-based Q-learning model in which a Q-learning-based reinforcement learning algorithm and a multi-agent framework are used to describe the complex activity–travel choice phenomena that characterize interactions among individuals in a mutually dynamic environment. Furthermore, reward functions were constructed based on traditional one-day-based OD survey data to increase the practical value of the model. The multi-agent-based Q-learning model was used to assess the influences of staggered working hours on individuals' lives by simulating changes in their activity–travel patterns. Based on the travel patterns of every participant, the time–space distribution of traffic in the transportation system was determined. The efficiency of this policy and the best approach to implementing it were evaluated in a rational and comprehensive way.

1. Modeling

1.1. Dynamic Environment

The geographic environment for the simulation was developed based on the Tongling city area. The target area covers 237 m^2 and is divided into 13 traffic analysis areas and 27 Traffic Analysis Zones (TAZs), according to land use conditions, as shown in Fig. 1. The model considers the major and collector roads between and within the TAZs.

The travel times between the TAZs was estimated using the road impedance function developed by the US Bureau of Public Roads (BPR). The travel times between the TAZs were ever-changing due to the trips generated. The road impedance function was originally developed to estimate vehicle travel times, but it can also be applied to other travel modes with recalibration of the parameters. However, the travel time associated with walking is considered only with respect to the travel distance.

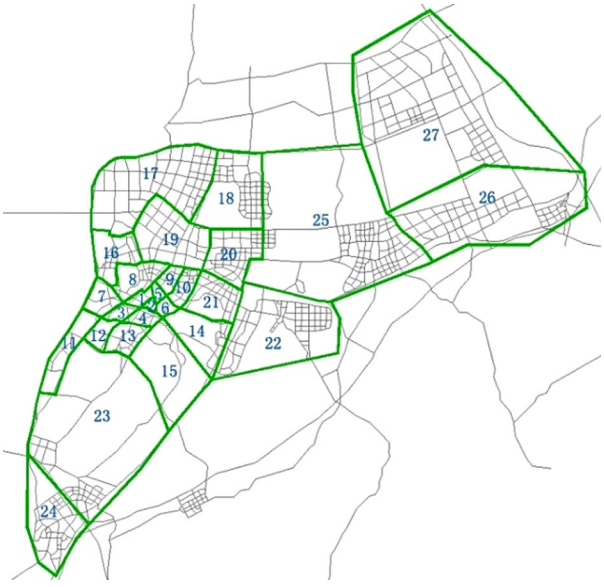

Fig. 1. Simulation environment

The mathematical expression of this model is as follows:

$$T_{nij}(V) = T_{0nij} \cdot \left(1 + \alpha \cdot \left(\frac{V_{ij}}{C_{ij}}\right)^{\beta}\right), \quad (1)$$

where: T_{nij} and T_{0nij} – the travel times of travel mode n when the volumes of traffic between traffic zones i and j are V and zero, respectively; V_{ij} and C_{ij} – The traffic volume and capacity between TAZ i and j; α and β – model parameters calibrated by survey data. Different travel modes are influenced by road impedance to different degrees, so the calibrated values of α and β for different travel modes are determined separately based on real-world data.

1.2. Agents Generation

In this study, trip makers are regarded as agents who make activity–travel decisions and influence other agents' decisions in a mutually dynamic environment. The agent initialization process generates heterogeneous individuals and households, each of which has a unique activity–travel pattern, based on real-world survey data.

OD survey data collected in Tongling in 2011 were used to establish various types of agents. The survey included collection of data on individual and household sociodemographics and travel records.

These travel records consist of departure and arrival times, origins and destinations, and travel modes and purposes. Trip purposes were divided into nine categories: going to work, going to school, official business, shopping, socializing-recreation, picking up, personal business, returning home and returning to work. Among these categories, going to work, going to school and official business were defined as commuting activities or simple work activities; shopping, picking up, and personal business were defined as maintenance activities; and socializing-recreation was defined as a leisure activity. Maintenance and leisure activities were further defined as non-working activities. Hence, the nine categories of activities could be divided into four types: work, maintenance activities, leisure activities and staying at home.

At the time of the OD survey, the population of Tongling was 392000. Usable travel data were obtained from 6676 residents who were more than 6 years old. Because students' activity–travel schedules are rather fixed and the main focus of this paper is on working and non-working groups, the students' data, which consisted of 2640 records, were not considered. Thus, 4036 data records were used to conduct the analysis described in this study. Of this total, 2726 records (68.0%) pertained to commuting activity patterns and 1310 records (32.0%) pertained to non-working activity patterns.

Because the classification of agents is based on activity–travel patterns, there should be a sufficiently large sample size for each activity pattern. Thus, twelve typical activity–travel patterns–five commuting patterns and seven non-working patterns were extracted from the survey data. These twelve typical activity–travel patterns correspond to twelve types of agents. Table 1 lists the

Table 1. Description of generated agents

Agent type	Description	Number	Ratio
hwh	Simple work pattern with only primary tour	1574	47.08%
hshwh	Simple work pattern as the primary tour, with a secondary tour	23	0.69%
hwhsh	Simple work pattern as the primary tour, with a secondary tour	33	0.99%
hwhwh	Work tour with home-based sub-tour	792	23.69%
hwswh	A sub-tour during work	32	0.96%
hsh	Simple maintenance tour	613	18.34%
hlh	Simple leisure tour	103	3.08%
hshlh	Both maintenance and leisure tours, with the prior one being a maintenance tour	37	1.11%
hshsh	Two maintenance tours	57	1.71%
hlhsh	Both maintenance and leisure tours, with the prior one being a leisure tour	22	0.66%
hssh	Two consecutive maintenance tours	35	1.05%
hlhlh	Two leisure tours	22	0.66%

descriptions of these agents. In the agent type codes, '*h*' represents 'staying at home,' '*w*' represents 'working,' '*s*' represents 'shopping' and '*l*' represents 'leisure activity.'

The 4036 agents generated from the survey results were extrapolated to a population of 392000. Apart from the surveyed 6676 people, we established activity–travel attribute data for 385324 agents using the Monte Carlo method.

1.3. Decision Making

Modeling individuals' activities or travel decision-making processes involves three steps: constructing reward functions, modeling cognitive learning for a single agent and modeling interaction among multiple agents. These three steps are described below:

Step 1 involves extraction of typical activity patterns from OD survey data and constructing initial reward functions for different agent groups.

Step 2 involves modeling an individual agent's cognitive learning behaviors when making activity–travel decisions based on a Q-learning algorithm, in which the agent's time and space choices can be considered an integrated unit.

Step 3 involves loading agents into the network to interact with each other in a multi-agent framework. Lastly, the temporal–spatial distribution of the urban traffic system as a whole is determined, and each individual's activity–travel schedule is revealed, recorded and analyzed.

1.3.1. Reward Functions

Rewards represent the immediate benefits that agents receive from the environment, but these benefits cannot be determined directly from the survey data. Hence, reward functions are applied to describe immediate rewards. The reward functions proposed in this research follow some basic assumptions put forward by other researchers, as follows:

- individuals derive a certain utility from allocating time to activities (Yamamoto, Kitamura 1999) and this utility depends on both the amount of time allocated and the time of day at which participation in the activity takes place (Ettema *et al.* 2004);
- individuals derive a certain disutility from the time spent travelling (Ben-Akiva, Lerman 1985);
- the utility of a discretionary activity is dependent on the activity history of the agent. In general, the longer ago that an agent last engaged in a certain activity, the greater the current utility of that activity will be (Arentze *et al.* 2010).

Four distinct reward functions were applied to represent the immediate rewards that agents receive from the environment in the reinforcement learning process.

The optimal data source for constructing reward functions is the survey data set from a multi-day GPS-based prompted-recall survey, which is used to capture the underlying activity attribute planning process. However, longitudinal travel surveys are costly and impose a high burden on respondents. A sufficiently large sample size is also necessary to allow segmentation with respect to spatial and sociodemographic variables. Because in China, one-day activity diary data are collected on a regular basis for multiple purposes, these data can be obtained at relatively low cost and provide the large sample sizes needed for forecasting and policy analysis purposes (Arentze *et al.* 2011). Therefore, we chose to utilize one-day activity diary data by subsuming individuals who share similar activity–travel patterns into several types and then calculating their activity–travel attributes, such as optimal start times for work, the average time spent on shopping or the most popular sites for entertainment.

Typical relationships between reward, activity start time and activity duration are shown in Fig. 2. It is assumed that if there are more people who prefer to engage in a certain activity at a certain start time for a certain length of time, the greater the reward will be for engaging in this activity at that start time and for that length of time. Using this approach, the rewards associated with individuals' activities and travel decisions were

extracted from calculated activity–travel attributes for the four reward functions described below:

a) Reward functions of activity duration

If the duration of a certain activity is within a reasonable range, it should result in a rather high cumulative reward for an agent. With increasing duration, fatigue effects come into play, resulting in a diminishing utility with increasing duration:

$$r_{duration}=\begin{cases}50\cdot d, \text{ when } d<d_{min};\\ 50\cdot d_{min}+60\cdot\left(d-d_{min}\right), \text{ when } d_{min}<d<d_{avg};\\ 50\cdot d_{min}+60\cdot\left(d_{avg}-d_{min}\right)-60\cdot\left(d-d_{avg}\right), \text{ when } d_{avg}<d<d_{max};\\ 50\cdot d_{min}+60\cdot\left(d_{avg}-d_{min}\right)-60\cdot\left(d_{max}-d_{avg}\right)-200\cdot\left(d-d_{max}\right), \text{ when } d_{max}<d.\end{cases} \quad (2)$$

where: d_{min}, d_{max} and d_{avg} – reasonable minimum, maximum and average durations, respectively, of an activity. These are the 5%, 95% and 50% percentile durations, respectively, of an activity, calculated based on the survey data. These durations are different for different activity patterns even if the action remains the same. For example, the d_{avg} for the first 'work' action in the '*hwhwh*' pattern is different from the d_{avg} for 'work' in the '*hwh*' pattern.

b) Reward functions of start time

Each activity's start time should be within a reasonable range. For example, for most people, going to work at 7:30 will yield a positive reward, while going to work at 2:00 will yield a negative reward. In this respect, it is assumed that there are some intrinsic preferences for the times of day at which certain activities are undertaken. We used polynomial functions to fit the distribution curve of reward as a function of start time:

$$r_{start\ time}=C_i\left(s\right), \quad (3)$$

where: C_i – the polynomial function of the ith action of a certain activity pattern; s – the start time of activity.

c) Reward functions of travel cost

Transferring from one activity to another often yields a negative reward. The reward associated with an individual trip T is defined as a relatively simple function of the travel time $R^E_{T,t}$ and the travel cost $R^E_{T,c}$ associated with trip T made in certain transportation environment E:

$$r_{travel}=\frac{-R^E_{T,t}-\left(R^E_{T,c}+R_{T,m}\right)}{VOT}, \quad (4)$$

where: E – the current transportation environment, mainly involving the degree of road congestion for every road section in the network (in this study, the transportation system is considered dynamic because interactions always exist among individuals; that is, the reward associated with an individual trip changes during the trip); $R_{T,m}$ – a constant that represents the constant utility of a trip made by mode m (for a bus trip, this value is equal to the ticket price, which was assumed be 2 Yuan in Tongling; for a car trip, this value is equal to the parking fee, which was assumed to be 20 Yuan in the downtown area and 10 Yuan in other areas of the city; for bike trips and walking trips, the value of $R_{T,m}$ is zero); $R^E_{T,c}$ – the travel cost for agents traveling by private cars, which is assumed to depend only on the travel distance, in this case, 2 Yuan per kilometre; VOT – the value of an agent's time, which is assumed to be related to a Family's Monthly Income (FMI), which is known from the OD survey data (VOT can be calculated as $VOT=\frac{FMI}{30\cdot 24\cdot 3600\cdot N}$, where N is the number of family members); $R^E_{T,t}$ – the time cost of trip T. The formula proposed in (Janssens *et al.* 2007) was used in this study to describe the reward function based on travel time and demarcate the parameters with real-world data according to a least squares algorithm:

$$R^E_{T,t}=c\cdot\left(b\cdot t\right)^a, \quad (5)$$

where: for walking trips: a = 1.4, b = 0.09, c = 5; for bike trips: a = 1.2, b = 0.11, c = 5; for car trips: a = 0.5, b = 0.22, c = 5; for public transit trips: a = 0.9, b = 0.14, c = 5. The term t represents the real travel time between two activities of each agent.

d) Reward functions of discretionary activity location attraction

In modeling individuals' destination choices for discretionary activities, consideration of the travel cost only is inadequate because it leads to a situation in which agents all choose the nearest shopping and entertainment destinations. However, it is a common phenomenon in reality that people travel far to entertainment centers that are more attractive. Therefore, the reward functions of location attraction for discretionary activities are constructed based on both individuals' preferences and the quality of the facilities in various traffic zones:

$$r_{attract}=att_i\left(A_j\right)\cdot F\left(l_i,A_j\right), \quad (6)$$

where: $F(l_i, A_j)$ is the basic utility function for conducting activity A_j (shopping or entertainment) in traffic zone l_i, which is related to the size and the level of service of a certain shopping or entertainment facility. Agglomeration effects for shopping malls and entertainment centers are also taken into consideration. The quality and quantities of shopping and entertainment facilities in certain traffic zones were extracted from land use data provided by the Urban Planning Bureau of Tongling.

att_i describes the reputation of a certain traffic zone with respect to activity A_j:

$$att_i\left(Aj\right)=\frac{n_i-n\min}{n_{\max}-n_{\min}}, \quad (7)$$

where: n_i – the number of activities A_j conducted in zone i; $n_{\min}$ and $n_{\max}$ – minimal and maximum numbers of activity A_j conducted among all zones.

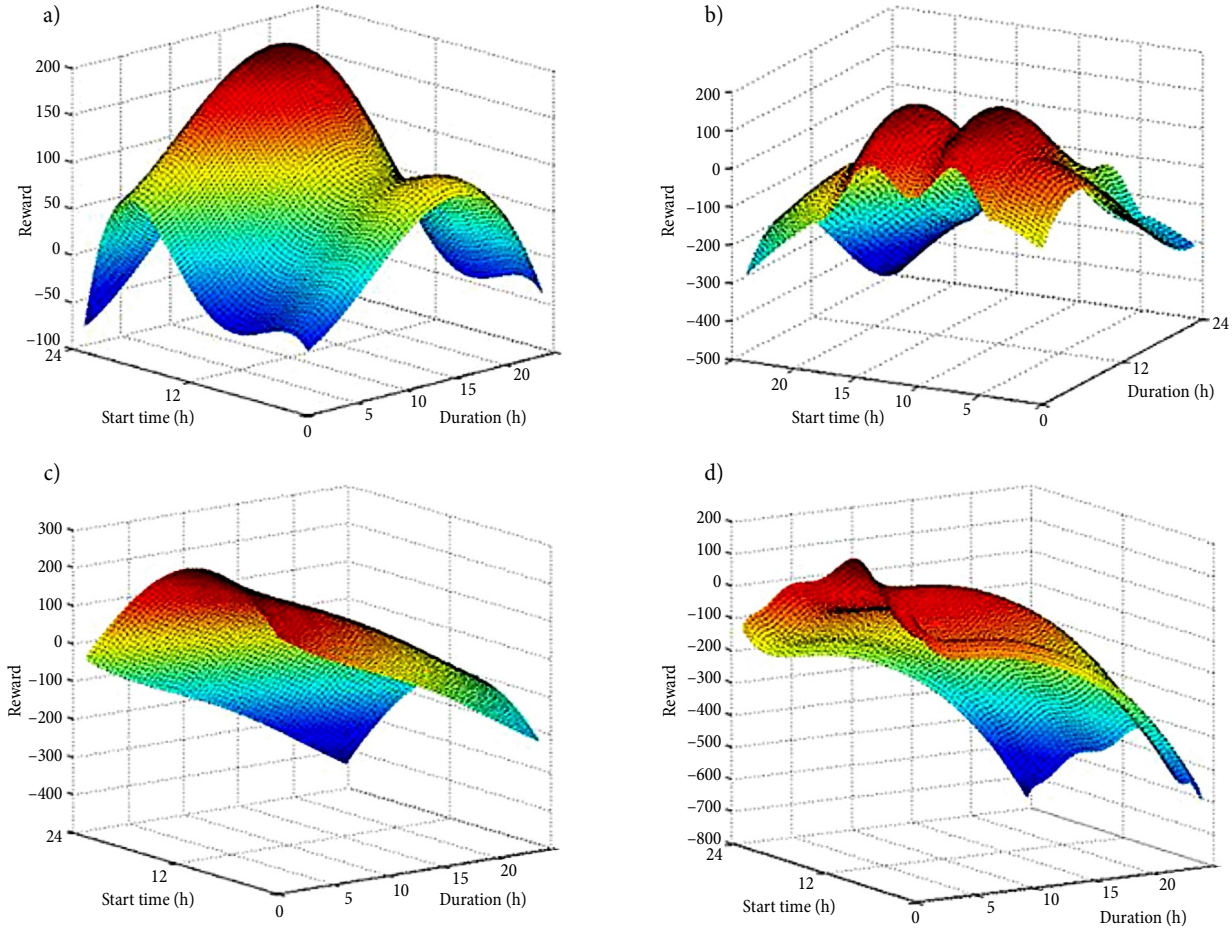

Fig. 2. Relationships between reward, activity start time and activity duration: a – home reward function; b – working reward function; c – leisure reward function; d – shopping reward function

1.3.2. Scheduling Activity–Travel Plan by Cognitive Learning

In this section, the Q-learning based reinforcement learning algorithm is introduced to describe the complex time–space choice behaviors of agents in the activity–travel plan scheduling process. Detailed examples of the Q-learning process and simulation results are presented.

Several basic concepts concerning the implementation of the Q-learning algorithm are defined below.

State: A vector of (activity, start time, duration, location and travel time) that represents an agent's state and is denoted by (a, s, d, l, t) for brevity.

Activity: Four activities were considered in the preliminary phase of this research: home, work, shopping and leisure. Shopping and leisure activities were considered to be types of discretionary activities.

Duration and start time: Time variables should be discrete in Q-learning. The unit time slot was 15 min, which divides a day into 96 time slots. Because the number of states should be finite, the longest duration of an activity was limited to 24 hours. Hence, both the duration and start time can be represented by numbers from 1 to 96.

Location: The location unit is a TAZ, an area that hosts multiple activities, including leisure, shopping and working, etc.

Action: An agent will randomly choose an action from an action choice set for every time slot. In general, there are two types of actions for every agent at every time step: continuing the current activity or changing to another activity.

Reward: The reward is defined as the immediate feedback that an action yields. In this study, the reward for an action is characterized by the degree of attraction of the location, the activity duration, the activity start time and the travel cost.

Q-Value: The Q-value is the total feedback that an action may yield in the short term or the long term.

Reinforcement learning tasks are generally treated in discrete time steps. At each time step t, the agent observes the current state and chooses a possible action to perform. The agent's subsequent state is $S_{t+1} = \delta(s_t, a_t)$, and the environment responds by giving the agent a reward: $r(s_t, a_t)$. It is probable that there is some delay associated with receiving preferred awards. For this reason, the task of the agent is to learn a policy $\pi : S \to A$, according to which the agent will receive the maximum

cumulative reward for one day. Given a random policy π from a random state S_t, the cumulative reward of S_t can be formulated as follows:

$$V^{\pi}(S_t) = r_t + \gamma r_{t+1} + \gamma^2 r_{t+2} + \cdots = \sum_{i=0}^{\infty} \gamma^i r_{t+i}, \quad (8)$$

where: r_{t+i} represents the scalar reward i steps after t; γ is the discounting factor. The agent only receives an immediate reward if γ is set to zero.

Obviously, the agent needs to learn the optimal policy $\pi^*(s)$ that maximizes the cumulative reward. Unfortunately, determining the optimum policy requires that knowledge of the immediate reward function r and the state transition function δ be known in advance, which is usually impossible in reality. That is, the domain knowledge is most likely not perfect. Q-learning serves to select optimal actions even when the agent has no knowledge about the reward and state functions.

We define $\widehat{Q}$ as the estimation of the true Q-value. The Q-learning algorithm maintains a large table with entries for each state–action pair. The Q-learning process can be described as follows:

1) The $\widehat{Q}(s,a)$ values are initialized with random numbers and stored.
2) A random starting state s is selected that has at least one possible action.
3) The agent observes its current state s and chooses a possible action a to perform, which leads to the next state. The immediate reward $r(s,a)$ and resulting new state $\delta(s_t, a_t)$ are determined.
4) The $\widehat{Q}(s,a)$ value of the state–action pair is updated according to the following rule:

$$\widehat{Q}(s,a) \leftarrow r(s,a) + \gamma \max_{a'} \widehat{Q}(s',a'). \quad (9)$$

5) Step 3 is repeated if the new state has at least one possible action. Otherwise, step 2 is repeated.

After the Q-values of state–action pairs have been well estimated by the Q-learning algorithm, the agent can reach a globally optimal solution by repeatedly selecting the actions that maximize the local values of Q for the agent's current state.

When implementing Q-learning in time–space choice simulation, mandatory activities, such as going to work and going to school, are considered to be fixed, while the locations for discretionary activities, such as maintenance and leisure activities, are flexible for agents. In addition, several constraints are identified that are consistent with common-sense notions:

- an agent's travel mode choices made during the course of a day must be in accordance with each other; for example, if an agent drives away from home, he or she must drive back home.
- public transit serves from 6:00 am to 10:00 pm.
- agents must get back home at or before 24:00.

Under these constraints, different agents may make different time–space choices based on their own attributions, established through the reinforcement learning process, to optimize their overall rewards. A Q-learning flow chart for four types of agents, '*hshwh*', '*hwhwh*', '*hwh*' and '*hwhsh*', is shown in Fig. 3. Taking the agent

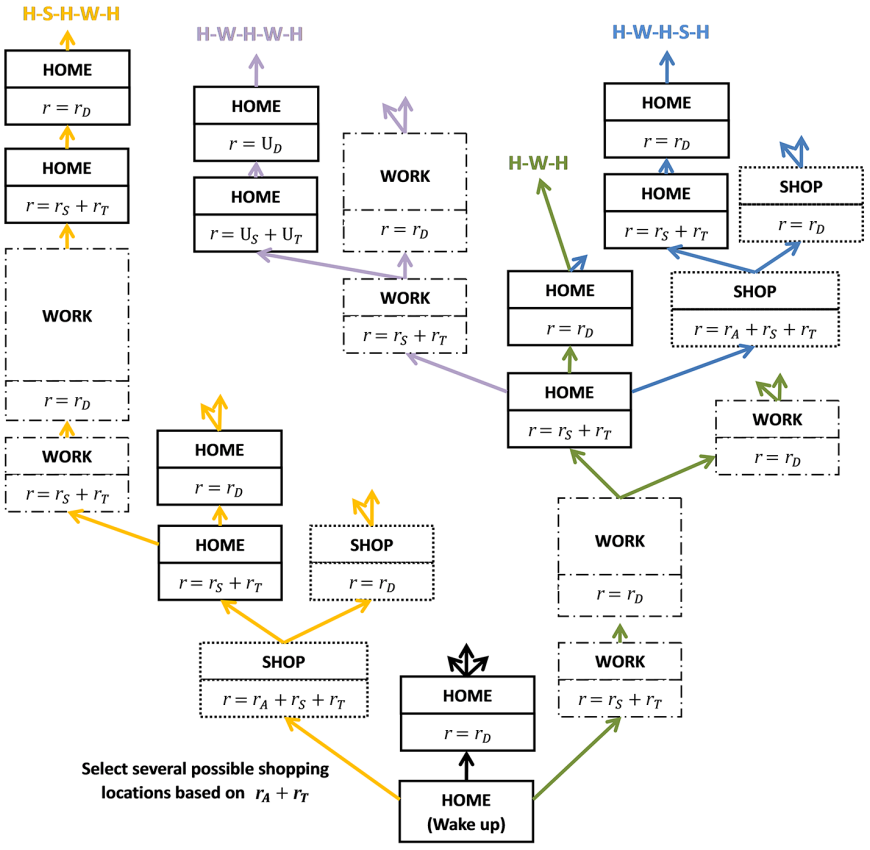

Fig. 3. Q-learning flow chart

'*hshwh*' as an example, the activity–travel plan scheduling process starts from the time the agent wakes up, after which the agent will randomly choose an action from a given choice set (in this case, {'Stay at Home', 'Go Shopping in traffic zone *x*1', 'Go Shopping in traffic zone *x*2' or 'Go Shopping in traffic zone *x*3'}. The agent then moves on with its action choice process, as shown in Fig. 3, under the given constraints. After each action is taken, the agent receives a reward *r*, which is the sum of the values of the corresponding reward functions. The corresponding *Q*-value in the *Q* matrix is then updated. Lastly, when the *Q* matrix achieves convergence, the agent's temporal–spatial choices in his or her activity–travel pattern can be simulated.

We randomly chose twelve individuals corresponding to twelve types of agents and simulated their time–space choice behaviors under fixed real-world traffic conditions. The simulation results are shown in Table 2. For example, 'Home (0:00,7:20) 6 <walk>…' means that the individual stays at home from 0:00 to 7:20 in the 6th traffic zone and then walks to the site of the next activity. The time–space choices of agents match reality without inconsistencies such as staying in an activity too long or traveling at an inappropriate time.

1.3.3. Considering Interaction Among Multiple Agents

In this study, a multi-agent framework was built in which all twelve types of individuals were defined as traveler agents and the dynamic environment was also regarded as an agent. The environment agent reacts to the activity–travel decisions of traveler agents by updating the degree of congestion for every road section, which in turns influences other traveler agents' decision-making behavior. For example, if a traveler agent moves from one zone to another, the volume of traffic between these two zones is updated according to the agent's departure time, arrival time and travel mode. Thus the environment, with a newly updated higher degree of congestion, provides a lower reward to the next traveler agent who initially decides to depart at the same time or to the same place. To achieve the maximum reward, the traveler agent may change his or her decision and select a new departure time or go to other places for the activity, which will again exert an influence on other agents. In this manner, the interaction among agents is simulated:

a) Temporal characteristics of simulation results

By taking interactions among individuals into consideration, all citizens' activity–travel schedules can be calculated, and from these schedules, the overall traffic distribution in Tongling can be determined. Fig. 4 shows a comparison of the temporal distributions of traffic flow from the survey data and from the simulation results obtained using the proposed multi-agent-based Q-learning model.

In an environment in which agent interactions are not considered, agents with the same properties fail to consider the limited traffic capacity and its influences on travelers' decisions, which results in abrupt changes in traffic flow and aberrant high volumes of traffic during peak hours. When agent interactions are considered, agents with the same properties make different decisions according to differences in the environment. This is consistent with reality: individuals may adjust their travel plans to avoid congestion. The accuracy of the

Table 2. Activity–travel schedule

Agent	Activity–travel schedule
'hwh'	Home (0:00,7:20) 6 <walk> Work (7:40,19:00) 2 <walk> Home (19:20,24:00) 6
'hsh'	Home (0:00,7:00) 6 <walk> Shop (7:50,8:30) 5 <walk> Home (8:35,24:00) 6
'hwhwh'	Home (0:00,7:40) 6 <bus> Work (7:50,11:30) 8 <bus> Home (11:50,13:40) 6 <bus> Work (3:50,17:45) 8 <bus> Home (18:0,24:00) 6
'hlh'	Home (0:00,6:00) 6 <walk> Leisure (7: 00,10: 00) 10 <walk> Home (11: 00,24: 00) 6
'hlhsh'	Home (0:00,6:00) 6 <walk> Leisure (6:10,7:30) 6 <walk> Home (7:40,8: 00) 6 <walk> Shop (8:10,9:30) 5 <walk> Home (9:40,24: 00) 6
'hshsh'	Home (0:00,7:50) 6 <walk> Shop (7:57,11:15) 6 <walk> Home (11:40,14:15) 6 <walk> Shop (14:25,17:40) 6 <walk> Home (18:50,24: 00) 6
'hwswh'	Home (0:00,7:50) 23 <bus> Work (7:50,8:10) 2 <car> Shop (8:20,17:10) 2 <car> Work (17:20,17:30) 2 <bus> Home (18:10,24:00) 23
'hwhsh'	Home (0:00,6:20) 12 <car> Work (7:45,17:15) 25 <car> Home (18:15,19:10) 12 <car> Shop (19:18,20:50) 2 <car> Home (20:15,24:00) 12
'hshwh'	Home (0:00,7:25) 5 <walk> Shop (7:55,8:50) 4 <walk> Home (9:20,16:20) 5 <bike> Work (16:30,22:10) 10 <bike> Home (22:20,24:00) 5
'hshlh'	Home (0:00,7:10) 15 <walk> Shop (7:30,8:55) 15 <walk> Home (9:20,16:10) 15 <walk> Leisure (16:20,18:0) 15 <walk> Home (18:50,24: 00) 15
'hssh'	Home (0:00,7:40) 6 <walk> Shop (8:0,8:05) 6 <walk> Shop (8:25,9:10) 15 <walk> Home (9:45,24:00) 6
'hlhlh'	Home (0:00,7:10) 8 <bike> Leisure (7:30,11:30) 10 <bike> Home (11:50,13:30) 8 <bike> Leisure (13:50,17:30) 10 <bike> Home (17:50,24:00) 8

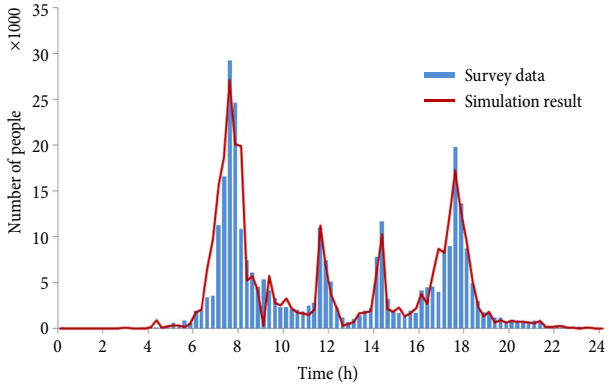

Fig. 4. Temporal flow of traffic distribution from survey data and from simulation using the multi-agent-based Q-learning model

simulation results demonstrates the advantage of multi-agent simulation. The correlation coefficient between the multi-agent simulation results and the survey data is 95%. In addition, traffic volumes during peak hours are important in traffic policy formulation. The relative standard error during the peak hours (7:00 to 8:00 and 17:30 to 18:30) between the multi-agent simulation results and the survey data is 12%.

b) Spatial characteristics of simulation results

It is reasonable to assume that every agent's living and working places are largely fixed and that only the zones for discretionary activities can be freely chosen. Fig. 5 and Fig. 6 show the simulation results for all agents' choices for daily shopping and leisure locations. Agents are most attracted to the zones with high degrees of attraction for shopping or entertaining, and they may then choose other zones that are also attractive for these elasticity trips because of traffic congestion.

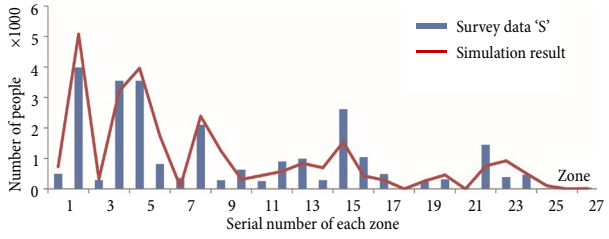

Fig. 5. Survey data and simulation results for shopping activity location choices using the multi-agent-based Q-learning model

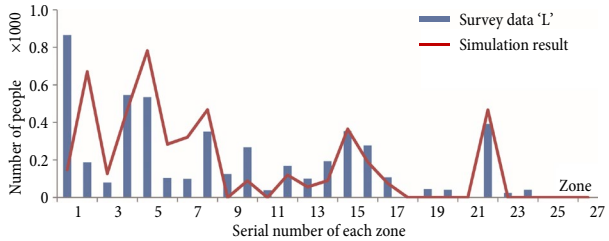

Fig. 6. Survey data and simulation results for leisure activity location choices using the multi-agent-based Q-learning model

Consideration of agent interaction yields more reasonable and realistic simulation results. Because zones with high degrees of attraction are very crowded, agents may choose to conduct their activities in other zones with lower degrees of attraction that are less congested. The correlation coefficient between the multi-agent simulation results and the survey data is 93% for shopping activity and 93% for leisure activity, which demonstrates the spatial accuracy of multi-agent simulation.

2. Evaluating Staggered Working Hours

As a TDM measure, the goal of staggering working hours is to reduce travel demand during peak hours through the adjustment of working hours. In China, the cities of Shenzhen, Chongqing and Hangzhou, among others, have long had policies in place to stagger working hours, and these policies have proven to be effective TDM measures. In China, local governments are able to adjust the work hours of government agencies and some public institutions, which makes staggering work hours practically feasible. The method proposed in this study, which is capable of reflecting the interactions of individual agents, overall traffic conditions in the network and the policy's impacts on individual activity scheduling, makes it possible to accurately assess the effect of staggered work hours through agent-based activity–travel pattern simulation.

2.1. Virtual Schemes for Staggered Working Hours

Staggered working hours had not been implemented in Tongling at the time that the OD survey was carried out. Thus, virtual schemes were considered for this case study, and the residents' activity–travel patterns were simulated in the agent-based activity scheduling model based on those schemes. Four virtual schemes were considered, involving postponing the start of work time at public institutions by 15 min, 30 min, 45 min or 1 h. The best scheme was then identified by analyzing the simulation results.

2.2. Simulation Results and Evaluation

2.2.1. Impacts on Individuals' Lives

Based on the simulation results, the 30-min scheme, in which people work from 8:30 a.m. to 6:00 p.m., was considered to be the best. The result shows that the traffic volumes during the morning and evening peak hours decreased significantly, and the original 15-min peak volume decreased by an average of 16%. Compared to this optimal scheme, the 15-min scheme had a limited effect on reducing the traffic volumes during the peak hours, with an average reduction of only 7%. The 45-min scheme and the 1-h scheme performed well in reducing traffic volume, with average reductions of 21% and 24%, respectively. However, these two schemes greatly disrupted residents' normal lives. For example, with the 45-min scheme, 91% of the residents who used to go shopping after work were able to allocate 0 or only 15 min to their shopping activities, which is the mini-

mum interval in the model. This implies that they might feel uncomfortable with the disruption to their lives that this scheme produces.

2.2.2. Impacts on the Overall Traffic System

The results show that a significant reduction in traffic volume from 7:00 to 8:00 resulted from a 30-min postponement in the working start time. The original 15-min peak volume decreased by 24%, as shown in Fig. 7.

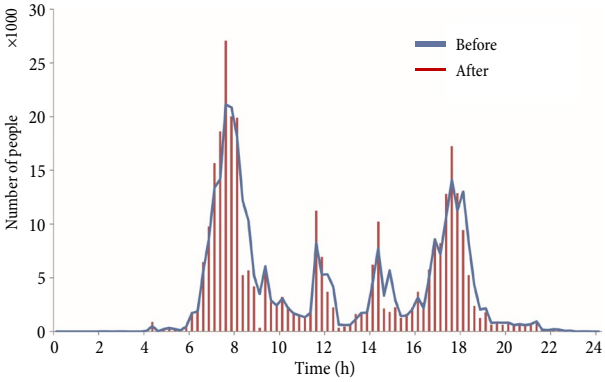

Fig. 7. Temporal distribution of traffic flow before and after the implementation of a policy of staggering working hours

To illustrate the spatial variation in traffic volumes that would result from the implementation of staggered working hours, several OD pairs were extracted, as shown in Table 3. It can be clearly observed that staggered work hour policy is an effective TDM measure for reducing the travel demand during the peak hours. The staggered work hour policy tends to reduce the travel demand more during the morning peak hour than during the evening peak hour. We believe that most travel demand in the morning peak is commuting travel, for which the working time determines the departure time. In contrast, during the evening peak, there is a considerable proportion of travel demand associated with elastic activities such as shopping and leisure activities. People will choose their departure times based on their activity choices, which diminishes the effect of the policy during the evening peak hour.

Table 3. Comparison of OD pairs before and after changing work start hours

OD pair	*Morning peak*			
	Before	After	Difference	Ratio
(6,2)	8826	7527	1299	14.7%
(15,2)	5406	4131	1275	23.6%
(8,8)	17322	15310	2012	11.6%
(5,10)	5658	4278	1380	24.4%
(7,11)	6447	4648	1799	27.9%
(13,13)	8120	6478	1642	20.2%
(15,15)	14311	12639	1672	11.7%
(16,16)	11260	9131	2129	18.9%
(22,22)	13554	11594	1960	14.5%
(23,23)	13501	10265	3236	24.0%
OD pair	*Evening peak*			
	Before	After	Difference	Ratio
(8,6)	4585	4088	497	10.8%
(2,8)	1936	1536	400	20.6%
(4,8)	2069	1738	331	16.0%
(8,8)	7393	7055	338	4.6%
(19,8)	4610	3852	758	16.4%
(10,10)	2393	2037	356	14.9%
(2,14)	2498	2139	359	14.4%
(2,15)	2516	2188	328	13.1%
(8,16)	2158	1635	523	24.3%
(16,16)	10737	10077	660	6.1%

Conclusions

The effects of staggered working hours on a traffic system were estimated by simulating individuals' daily activity–travel patterns using a multi-agent-based Q-learning model. The study and its main findings are summarized below:

- The effects that have been taken into account include the influences on individuals' daily activity–travel schedules and the traffic system. The simulation model used shows how staggered working hours affects the traffic volume during peak hours and balances the spatial distribution of traffic. The proposed model can be used to evaluate other TDM policies using traditional survey data.
- A multi-agent-based Q-learning model was proposed in this study to simulate individuals' activity–travel scheduling behavior. Reward functions were constructed based on a traditional one-day-based OD survey. Individuals' time and space choices with respect to activity–travel patterns were simulated simultaneously using a Q-learning algorithm. Interactions among individuals were taken into account by establishing a mutually dynamic environment.

The following are our ongoing work:

- In this study, the activity pattern of each agent was fixed. In future research, we could make agents decide their patterns dynamically by establishing multi-category reward functions based on panel data for one week or more.
- Research has shown that the greatest potential of TDM lies in the integration of TDM strategies. The multi-agent-based Q-learning model, which has strong compatibility and expandability, can be used to simulate complex individual behavior under integrated TDM strategies, evaluate the influences of TDM strategies on a traffic system and reveal how different TDM strategies complement and reinforce each other, by changing the values of the parameters in the reward functions.

Acknowledgments

This research was supported by the National Basic Research 973 Program (2012CB725400) and the National Natural Science Foundation of China (51378120, 51338003, and 50908052). Fundamental Research Funds for the Central Universities and the Foundation for Young Key Teachers of Southeast University are also appreciated.

References

Adler, J. L.; Satapathy, G.; Manikonda, V.; Bowles, B.; Blue, V. J. 2005. A multi-agent approach to cooperative traffic management and route guidance, *Transportation Research Part B: Methodological* 39(4): 297–318. http://dx.doi.org/10.1016/j.trb.2004.03.005

Arentze, T. A.; Ettema, D.; Timmermans, H. J. P. 2011. Estimating a model of dynamic activity generation based on one-day observations: method and results, *Transportation Research Part B: Methodological* 45(2): 447–460. http://dx.doi.org/10.1016/j.trb.2010.07.005

Arentze, T. A.; Ettema, D.; Timmermans, H. J. P. 2010. Incorporating time and income constraints in dynamic agent-based models of activity generation and time use: approach and illustration, *Transportation Research Part C: Emerging Technologies* 18(1): 71–83. http://dx.doi.org/10.1016/j.trc.2009.04.016

Arentze, T.; Timmermans, H. 2008. Social networks, social interactions, and activity-travel behavior: a framework for microsimulation, *Environment and Planning B: Planning and Design* 35(6): 1012–1027. http://dx.doi.org/10.1068/b3319t

Arentze, T. A.; Timmermans, H. J. P. 2004. A learning-based transportation oriented simulation system, *Transportation Research Part B: Methodological* 38(7): 613–633. http://dx.doi.org/10.1016/j.trb.2002.10.001

Bellemans, T.; Bothe, S.; Cho, S.; Giannotti, F.; Janssens, D.; Knapen, L.; Körner, C.; May, M.; Nanni, M.; Pedreschi, D.; Stange, H.; Trasarti, R.; Yasar, A.-U.-H.; Wets, G. 2012. An agent-based model to evaluate carpooling at large manufacturing plants, *Procedia Computer Science* 10: 1221–1227. http://dx.doi.org/10.1016/j.procs.2012.08.001

Ben-Akiva, M. E.; Lerman, S. R. 1985. *Discrete Choice Analysis: Theory and Application to Travel Demand*. The MIT Press. 384 p.

Benenson, I.; Martens, K.; Birfir, S. 2008. PARKAGENT: an agent-based model of parking in the city, *Computers, Environment and Urban Systems* 32(6): 431–439. http://dx.doi.org/10.1016/j.compenvurbsys.2008.09.011

Bhat, C. R.; Guo, J. Y.; Srinivasan, S.; Sivakumar, A. 2004. Comprehensive econometric microsimulator for daily activity-travel patterns, *Journal of Transportation Research Board* 1894: 57–66. http://dx.doi.org/10.3141/1894-07

Charypar, D.; Nagel, K. 2005. Q-learning for flexible learning of daily activity plans, *Transportation Research Record* 1935: 163–169. http://dx.doi.org/10.3141/1935-19

Cherchi, E.; Polak, J. W. 2005. Assessing user benefits with discrete choice models: implications of specification errors under random taste heterogeneity, *Transportation Research Record* 1926: 61–69. http://dx.doi.org/10.3141/1926-08

Chorus, C. G.; Timmermans, H. J. P. 2009. Measuring user benefits of changes in the transport system when traveler awareness is limited, *Transportation Research Part A: Policy and Practice* 43(5): 536–547. http://dx.doi.org/10.1016/j.tra.2009.02.002

Creutzig, F.; He, D. 2009. Climate change mitigation and co-benefits of feasible transport demand policies in Beijing, *Transportation Research Part D: Transport and Environment* 14(2): 120–131. http://dx.doi.org/10.1016/j.trd.2008.11.007

D'Este, G. 1985. The effect of staggered working hours on commuter trip durations, *Transportation Research Part A: General* 19(2): 109–117. http://dx.doi.org/10.1016/0191-2607(85)90021-4

Dia, H. 2002. An agent-based approach to modelling driver route choice behaviour under the influence of real-time information, *Transportation Research Part C: Emerging Technologies* 10(5–6): 331–349. http://dx.doi.org/10.1016/S0968-090X(02)00025-6

Dong, X.; Ben-Akiva, M. E.; Bowman, J. L.; Walker, J. L. 2006. Moving from trip-based to activity-based measures of accessibility, *Transportation Research Part A: Policy and Practice* 40(2): 163–180. http://dx.doi.org/10.1016/j.tra.2005.05.002

Ettema, D.; Ashiru, O.; Polak, J. W. 2004. Modeling timing and duration of activities and trips in response to road-pricing policies, *Transportation Research Record* 1894: 1–10. http://dx.doi.org/10.3141/1894-01

Ettema, D.; Bastin, F.; Polak, J.; Ashiru, O. 2007. Modelling the joint choice of activity timing and duration, *Transportation Research Part A: Policy and Practice* 41(9): 827–841. http://dx.doi.org/10.1016/j.tra.2007.03.001

Gutiérrez-i-Puigarnau, E.; Van Ommeren, J. N. 2012. Start time and worker compensation: implications for staggered-hours programmes, *Journal of Transport Economics and Policy* 46(2): 205–220.

Holmgren, J.; Davidsson, P.; Persson, J. A.; Ramstedt, L. 2012. TAPAS: A multi-agent-based model for simulation of transport chains, *Simulation Modelling Practice and Theory* 23: 1–18. http://dx.doi.org/10.1016/j.simpat.2011.12.011

Hug, K.; Mock-Hecker, R.; Würtenberger, J. 1997. Transport demand management by electronic fee collection in a zone-based pricing scheme: the Stuttgart MobilPASS field trial, *Transportation Research Record* 1576: 67–76. http://dx.doi.org/10.3141/1576-09

Hunt, J. D.; Stefan, K. J.; Brownlee, A. T.; Sun, S.; Basu, D. 2012. Short distance personal travel model (SDPTM) in the California statewide transportation demand model, in *TRB 91st Annual Meeting Compendium of Papers DVD*, 22–26 January 2012, Washington DC. 20 p. (DVD).

Janssens, D.; Lan, Y.; Wets, G.; Chen, G. 2007. Allocating time and location information to activity–travel patterns through reinforcement learning, *Knowledge-Based Systems* 20(5): 466–477. http://dx.doi.org/10.1016/j.knosys.2007.01.008

Mun, S.-I.; Yonekawa, M. 2006. Flextime, traffic congestion and urban productivity, *Journal of Transport Economics and Policy* 40(3): 329–358.

Timmermans, H.; Arentze, T.; Joh, C.-H. 2002. Analysing space-time behaviour: new approaches to old problems, *Progress in Human Geography* 26(2): 175–190. http://dx.doi.org/10.1191/0309132502ph363ra

Yamamoto, T.; Kitamura, R. 1999. An analysis of time allocation to in-home and out-of-home discretionary activities across working days and non- working days, *Transportation* 26(2): 231–250. http://dx.doi.org/10.1023/a:1005167311075

Yang, M.; Wang, W.; Chen, X. W.; Wan, T.; Xu, R. 2007. Empirical analysis of commute trip chaining: case study of Shangyu, China, *Transportation Research Record* 2038: 139–147. http://dx.doi.org/10.3141/2038-18