

FANG ZONG, Ph.D.

E-mail: zongfang@jlu.edu.cn

JIA HONGFEI, Ph.D.

E-mail: jiahf@jlu.edu.cn

Jilin University, College of Transportation
5988 Renmin Street, Changchun, Jilin, 130022, China

PAN XIANG, Ph.D.

E-mail: panx@zjut.edu.cn

Zhejiang University of Technology
866 Yuhangtang Road, Hangzhou, Zhejiang Province,
310058, China

WU YANG, M.S.

E-mail: jlu.zf@163.com

Jilin University, College of Transportation
5988 Renmin Street, Changchun, Jilin, 130022, China

Science in Traffic and Transport

Preliminary Communication

Accepted: Nov. 22, 2012

Approved: Sep. 24, 2013

PREDICTION OF COMMUTERS' DAILY TIME ALLOCATION

ABSTRACT

This paper presents a model system to predict the time allocation in commuters' daily activity-travel pattern. The departure time and the arrival time are estimated with Ordered Probit model and Support Vector Regression is introduced for travel time and activity duration prediction. Applied in a real-world time allocation prediction experiment, the model system shows a satisfactory level of prediction accuracy. This study provides useful insights into commuters' activity-travel time allocation decision by identifying the important influences, and the results are readily applied to a wide range of transportation practice, such as travel information system, by providing reliable forecast for variations in travel demand over time. By introducing the Support Vector Regression, it also makes a methodological contribution in enhancing prediction accuracy of travel time and activity duration prediction.

KEYWORDS

time allocation, commuting, activity, travel, Support Vector Regression

1. INTRODUCTION

An individual's daily activity-travel schedule is an important determinant of the temporal pattern of traffic demand. Understanding time allocation behaviour is not only indispensable in the development of full-scale model of daily activity-travel patterns but also crucial in forecasting temporal variation in travel demands and informing time-dependent transportation policies and measures, such as flexible working hours and real-time traveller information system [1, 2].

A daily activity-travel schedule includes all the timing and duration of all the activities and trips in a day. Since individuals have limited time, they need to make joint decisions of the timing and duration of all the trips in a day. However, timing and duration have been largely treated separately in previous literature, such as the report in [3], and many existing studies only considered part of daily activity agenda, like the one reported in [4]. Although the number of studies on daily activity-travel modelling, which covers both timing and duration prediction, has seen considerable increase in recent years, the accuracy of time prediction is determined by the performance of a whole model system, which usually includes activity pattern model and mode choice model besides time allocation model as well. It is argued that an *ad hoc* time allocation model may enhance the accuracy of timing and duration prediction.

This study is aimed at addressing the above issues by developing a model system to predict the typical daily commute activity-travel schedule, which includes both timing and duration of daily activities and trips. The reason for modelling commute time allocation is that commute trips are one of the major causes of congestion during the peak hours, e.g., according to the travel survey data in Beijing (2005), over 32% of the total trips during peak hours are commute trips [5].

The remainder of this paper is organized as follows. In Section 2 a review on activity-travel time allocation literature in general is presented. The data and the conceptual framework of the model are described in Section 3. This is followed by the model of travel departure time and stop arrival time in Section 4 and the

model of travel time and activity duration in Section 5. Prediction of a commute activity-travel schedule is presented in Section 6. The paper closes with some major conclusions and a discussion of future research directions.

2. EXISTING LITERATURE

A number of studies investigated the timing for a wide span of activities. For example, Bowman and Ben-Akiva [6] predicted the departure time of individual's daily tours by using Multinomial logit (MNL) model. Bhat [7] modelled the departure time of shopping trip using an Ordered generalized extreme value (OGEV) model. Small [8] constructed an MNL time allocation model for home-to-work morning departure time choice. Small [9] compared MNL, OGEV and Nested logit model in time of day modelling. As for duration prediction, Bhat and Steed [10] estimated a Hazard model of urban shopping trip durations. Juan and Xianyu [11] predicted daily travel time using the Hazard model.

Instead of analyzing the activity timing and duration separately, joint analysis can capture the interdependence between timing and duration. Pendyala and Bhat [12] used a joint discrete-continuous model to investigate the relationship between maintenance activity start time and activity duration. Other examples, to name a few, are one report in [13] on work start time and work duration, and another study reported in [14] on relationship between commuting time and

work duration, etc. Although these studies estimated timing and duration jointly, they were concerned about only part of the daily activity-travel pattern.

In recent years, several travel demand model systems have been developed to predict the entire daily activity-travel pattern. For instance, Bowman and Ben-Akiva [6] developed an activity-based travel demand model system, which consisted of daily activity pattern model, time allocation model, and mode choice model. Kitamura et al. [15] presented a sequential simulation approach to model daily activity-travel pattern, which includes the type, duration, location and mode of daily activities and trips. Guo and Bhat [16] constructed a model system to predict both workers' and non-workers' activity-travel patterns, including departure time, activity duration, destination and mode choice. Since these travel demand model systems jointly model time decision, mode choice and activity pattern, the accuracy of time prediction is determined by the performance of the whole model system. Therefore, instead of modelling the daily activity pattern, travel mode and time allocation with one model system, a separate daily activity-travel time prediction model is developed in this paper.

3. DATA AND MODELLING FRAMEWORK

3.1 Data

This study employs data from a large-scale daily travel survey conducted in Beijing in 2005 [5]. The sur-

Table 1 - Variables and statistics based on survey data

Variable	Level	Percentage (%)	Variable	Level	Percentage (%)
Income	1500RMB ¹	13.46	Gender	Male	50.96
	1501-2500RMB	24.90		Female	49.04
	2501-3500RMB	24.18	Occupation	Blue-collar worker (manufacture, construction, maintenance, etc.)	42.44
	3501-5500RMB	23.92		Administration	3.48
	5501-10000RMB	11.86		Education	30.29
	10001-20000RMB	1.41		Services	9.19
	20001-30000RMB	0.16		Health care	2.49
	Over 30001	0.11		Other	12.11
Trip purpose	Work ²	81.80	Mode	Walking	46.52
	Maintenance	7.87		Bike	48.03
	Entertainment	0.97		Bus	0.34
	Other	9.37		Auto	5.11
Variable	Mean	Standard deviation	Variable	Mean	Standard deviation
Age	32.93	15.32	Distance (km)	2.81	3.42

¹ RMB (Renminbi, also CNY or CN¥) is the official currency of China with 1RMB being equal to about \$0.16;

² It refers to going to school for students.

vey area covered 16,410.54 square kilometres (covering all the 18 districts of Beijing), and the residing population of the study was about 11.8 million of people. The survey sampled approximately 1.17% of the residing population. In addition to weekday OD information, it also collected information regarding household (size, car ownership, home location, monthly income and mobility), individuals (age, gender, occupation and driving license) and trips (departure time, arrival time, purpose, mode sequence, and transit path, etc.). With observations with missing values eliminated, the final sample consists of 202,883 trip records of 138,480 commuters from 54,398 households.

Table 1 presents all the socio-demographic and trip characteristic variables in the model, which have proven to be important influences in time allocation decisions in previous studies [3, 7, 17].

3.2 Analysis of the commute activity-travel agenda

In general, the daily commute activity-travel pattern (shown in Figure 1) is characterized by five different (sub-)patterns:

- Before-work pattern, which represents the activity-travel undertaken before leaving home to work in the morning;
- Home-to-work commute pattern, which represents the activity-travel undertaken during the morning commutes;
- Work-based sub-pattern, which represents the activity-travel undertaken from work during daytime;
- Work-to-home commute pattern, which represents the activity-travel undertaken during the evening commutes;
- Post-work pattern, which represents the activity-travel undertaken after arriving home at the end of the evening commute.

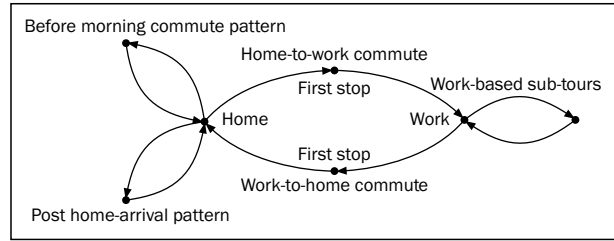


Figure 1 - Diagrammatic representation of commuters' daily activity-travel pattern

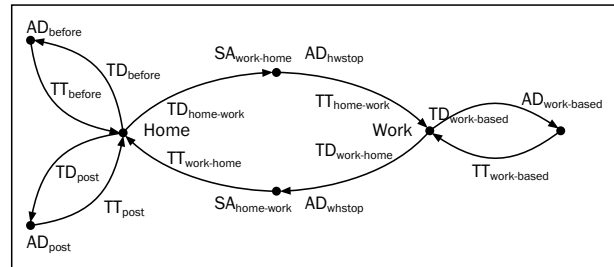


Figure 2 - Key time and duration values in daily activity-travel pattern

There may be more than one tour and also stops in each pattern. However, in order to reduce the complexity of the model, only the first tour in each pattern and only the first stop in the home-to-work / work-to-home commute pattern are modelled in this paper.

As shown in Figure 2 the key time and duration values in daily activity-travel pattern include:

- Travel departure time: departure time of before-work tour (TD_{before}), departure time of home-to-work commute trip ($TD_{home-work}$), departure time of work-based sub-tour ($TD_{work-based}$), departure time of work-to-home commute trip ($TD_{work-home}$), and departure time of post-work tour (TD_{post}).
- Stop arrival time: arrival time of home-to-work first stop ($SA_{home-work}$) and that of work-to-home first stop ($SA_{work-home}$).

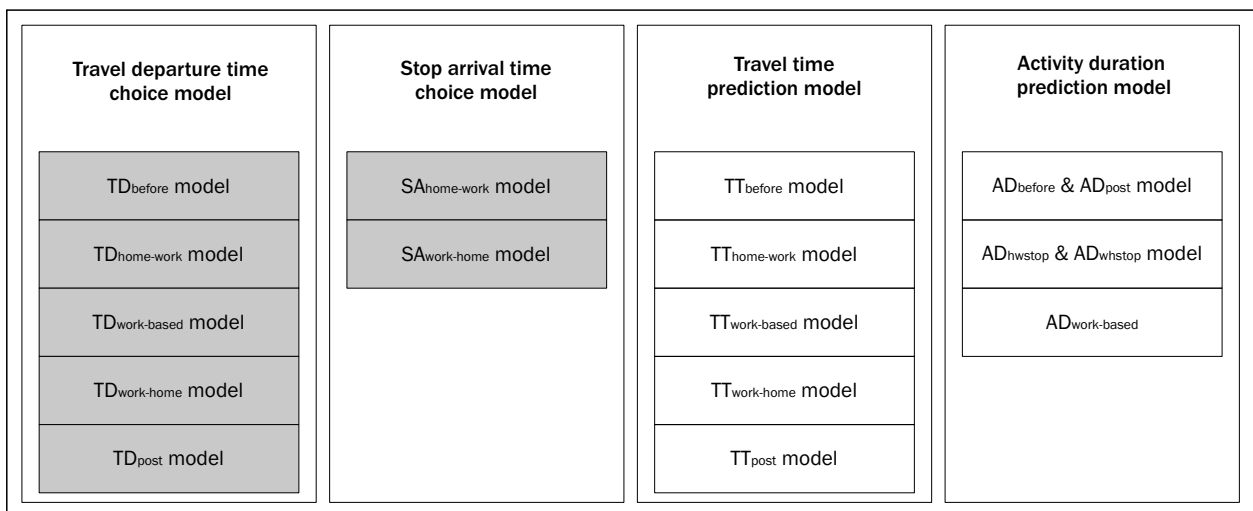


Figure 3 - Commuters' daily schedule modelling framework

Table 2 - Alternatives in the departure time and start time choice models

Model		Alternatives			
		1	2	3	4
Travel departure time choice models	TD _{before}	[3:00,6:00]	(6:00,7:30]	(7:30,9:00]	(9:00,10:00]
	TD _{home-work}	[4:00,6:00]	(6:00,7:30]	(7:30,9:00]	(9:00,11:00]
	TD _{work-based}	[9:00,11:00]	(11:00,13:00]	(13:00,16:00]	-
	TD _{work-home}	[12:00,16:00]	(16:00,17:30]	(17:30,19:00]	(19:00,21:00]
	TD _{post}	[18:00,19:00]	(19:00,20:00]	(20:00,23:00]	-
Stop arrival time choice models	SA _{home-work}	[5:00,7:00]	(7:00,8:00]	(8:00,9:00]	(9:00,11:30]
	SA _{work-home}	[13:00,16:30]	(16:30,18:00]	(18:00,19:30]	(19:30,21:30]

- (3) Travel time: before-work travel time (TT_{before}), home-to-work travel time (TT_{home-work}), travel time of work-based sub-trip (TT_{work-based}), work-to-home travel time (TT_{work-home}), and post-work travel time (TT_{post}).
- (4) Activity duration: before-work activity duration (AD_{before}), home-to-work stop duration (AD_{hwstop}), work-based sub-activity duration (AD_{work-based}), work-to-home stop duration (AD_{whstop}), and post-work activity duration (AD_{post}). The work (school) stay duration is not discussed in this paper, since generally the daily work time and study time is eight hours for most of the workers and students in China.

Then, the modelling framework is set up, which is composed of four categories of models as shown in Figure 3.

4. TRAVEL DEPARTURE TIME AND STOP ARRIVAL TIME MODELLING

4.1 Alternatives in the models

In order to model the travel departure time and stop arrival time, the continuous time was divided into discrete time intervals. These alternative time intervals are chosen by analyzing the distributions of travel departure times and stop arrival times based on the survey data. Moreover, in order to analyze and predict the pattern of time allocation in the peak periods, two specific alternatives were set, in the morning peak hours (6:00 am- 9:00 am) and two alternatives in the evening peak hours (16:00 pm- 19:00 pm) for TD_{home-work} and TD_{work-home} model, respectively. Table 2 shows the alternatives for five departure time choice models and two arrival time choice models.

4.2 Ordered Probit model

As shown in Table 2, the alternatives in travel departure time and stop arrival time choice models are ordered time periods. Since MNL model, which is commonly used in discrete choice modelling, would fail to account for the ordinal nature of the dependent variable and have the problem of IIA (Independence from

irrelevant alternatives) [18], this study will employ Ordered multiple choice model for departure time and arrival time modelling.

The Ordered multiple choice model assumes the relationship:

$$\sum_{j=1}^J P_n(j) = F(\alpha_j - \beta_j X_n, \theta), j = 1, \dots, J - 1 \quad (1)$$

and

$$P_n(J) = 1 - \sum_{j=1}^{J-1} P_n(j) \quad (2)$$

where $P_n(j)$ is the probability that alternative j is chosen as departure time of trip n ($n = 1, \dots, N$), α_j is an alternative specific constant, X_n is a vector of the attributes of trip n , β_j is a vector of estimable coefficients, and θ is a parameter that controls the shape of probability distribution F . Therefore, F can have various shapes of distribution based on different value of θ .

The Ordered Probit model, which assumes standard normal distribution for F is the most commonly used Ordered multiple choice model [19]. The Ordered Probit model has the following form:

$$P_n(1) = \Phi(\alpha_1 - \beta_1 X_n)$$

$$P_n(j) = \Phi(\alpha_j - \beta_j X_n) - \Phi(\alpha_{j-1} - \beta_{j-1} X_n), j = 2, \dots, J - 1 \quad (3)$$

$$P_n(J) = 1 - \sum_{j=1}^{J-1} P_n(j)$$

where $P_n(j)$ is the cumulative standard normal distribution function. For all the probabilities to be positive, we must have $\alpha_1 < \alpha_2 < \dots < \alpha_{J-1}$.

4.3 Estimation results

By using `oprobit()` function in Stata [20], both the travel departure time and the stop arrival time choice models are estimated and the results are shown in Table 3.

The results indicate that the older the commuter, the earlier the departure time of before-work trip, while old people tend to depart later for both home-to-work and work-to-home commute trips. Moreover, older commuters are also found to make home-to-work first stop later than younger commuters. Gender

Table 3 - Estimation results of the departure time and arrival time choice models

Variables	Travel departure time choice models										Stop arrival time choice models			
	TD _{before}		TD _{home-work}		TD _{work-based}		TD _{work-home}		TD _{post}		SA _{home-work}		SA _{work-home}	
	Coef.	Z-stat.	Coef.	Z-stat.	Coef.	Z-stat.	Coef.	Z-stat.	Coef.	Z-stat.	Coef.	Z-stat.	Coef.	Z-stat.
Gender	0.17	1.62	0.09	5.17	0.10	1.03	-0.06	-3.75	-0.27	-2.86	0.06	2.64	-	-
Age	-0.03	-5.43	0.02	26.83	-. ¹	-	0.01	9.03	-	-	0.01	9.94	-	-
Occupation	-	-	-0.02	-4.17	-0.06	-2.15	-	-	-	-	-0.02	-2.59	-0.01	-1.64
Income	-0.04	-1.00	0.08	12.78	0.07	1.78	0.04	6.62	-	-	0.12	12.72	0.05	5.66
Mode	0.16	1.95	-0.11	-10.40	-0.34	-5.89	0.11	10.72	-0.15	-2.14	-	-	-	-
Distance	-0.13	-2.36	-0.03	-9.92	0.05	2.48	0.02	7.69	0.09	2.29	-	-	0.02	1.41
Trip purpose	-0.34	-3.58	-	-	0.30	4.42	-	-	0.14	1.72	-0.14	-2.88	0.23	7.38
α_1	-3.31	-	-1.25	-	-0.22	-	-0.51	-	0.04	-	-0.14	-	0.01	-
α_2	-1.30	-	1.05	-	1.55	-	1.12	-	1.48	-	1.09	-	1.70	-
α_3	-0.09	-	2.72	-	-	-	2.02	-	-	-	1.97	-	2.48	-
P value	0.0000		0.0000		0.0000		0.0000		0.0000		0.0000		0.0000	
N	517		18,615		612		19,032		638		8,751		9,726	
Hit ratio (%)	60.54		64.52		66.01		56.96		57.21		49.08		59.35	

¹ All the “-” in this table denote that the independent variable is insignificant in the model.

affects the decisions of all the daily travel departure times and stop arrival times except stop arrival time of work-to-home pattern. Compared with women, men tend to depart later for work-to-home commute trip and post-work evening trip, while earlier for the other four types of trips (stops). In addition, the travel departure times and stop arrival times get later for commuters with higher income, except before-work departure time, which moves up with the increase in commuters' income. According to the survey data, people with higher income are older in general, and older people are more likely to leave home earlier for before-work morning activities, e.g., morning exercise. On the contrary, people with higher income are more flexible in terms of working schedule than those with lower income, which makes them have the luxury of leaving home for work later.

As for the trip-related independent variables, travel distance is proven to be significant for all the models except SA_{home-work} model. For before-work morning trip and home-to-work commute trip, the longer the travel distance, the earlier the commuter leaves home. On the contrary, the longer the distance, the later the work-based travel departure time, work-to-home travel departure time, post-work travel departure time and stop arrival time of work-to-home pattern are. In addition, mode does not affect the decision on stop arrival time, but has impact on choice of travel departure time. The reason is that it is mainly the activity of the stop which leads to the stop that decides the stop arrival time, while it is not only the activity but also the travel time that affects the travel departure time. Furthermore, since the mode has in-

fluence on travel time, then it has impact on travel departure time.

The results also show that the higher the speed of the mode, the later the before-work travel departure and work-to-home travel departure times are; while the higher the speed, the earlier the home-to-work travel departure time, work-based travel departure time and post-work travel departure times are.

5. TRAVEL TIME AND ACTIVITY DURATION MODELLING

5.1 SVR model

As one of the Support Vector Machines (SVM)-machine learning methods analyzing data and recognizing patterns, the Support Vector Regression (SVR) is widely applied in building a regression model with continuous dependent variable [21, 22]. We made the first effort of introducing SVR in estimating the travel time and activity duration.

Given a set of input-output data pairs $(x_1, y_1), (x_2, y_2), \dots, (x_l, y_l)$ ($x_i \in X \subseteq R^m, y_i \in Y \subseteq R^n, l$ is the number of training samples), that are randomly and independently generated from an unknown function, SVM estimates the function using the following Equation [23]:

$$f(x) = w \cdot \Phi(x) + b \quad w, x \in R^m, b \in R^n \quad (4)$$

where $\Phi(x)$ represents the high-dimensional feature spaces which are non-linearly mapped from the input space x . Here w denotes a parameter vector and b is the threshold [24]. If the domain of output space y contains continuous real values, the learning problem

then refers to SVR. Otherwise, if the interpretation y only takes category values, i.e., -1 and +1, it denotes Support Vector Classification (SVC) [25].

For SVR, the coefficients w and b in Equation (4) can be estimated by the so-called regularized risk functional:

$$MinJ = \frac{1}{2} \|w\|^2 + C \cdot R_{emp}[f] \tag{5}$$

The first term $1/2 \|w\|^2$ is called the regularized term which is used as the measurement of function flatness. The second term $R_{emp}[f]$ is the so-called loss function to measure the empirical error. C is regularization constant to determine the trade-off between the training error and the generalization performance. Here, the ε -insensitive loss function is employed to measure the empirical error:

$$|y - f(x)|_{\varepsilon} = \max\{0, |y - f(x)| - \varepsilon\} \tag{6}$$

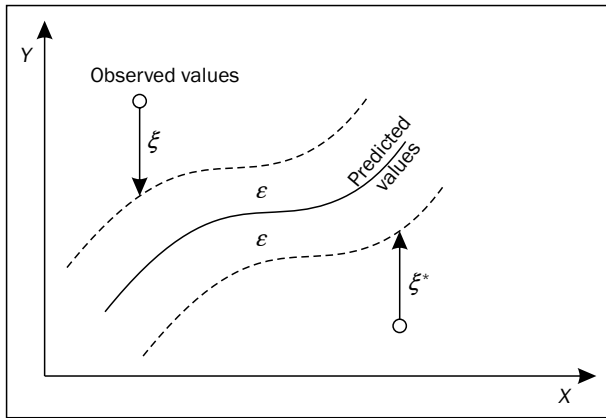


Figure 4 - Parameters for Support Vector Regression

Equation (6) defines ε tube (shown in Figure 4). The loss is zero if the predicted value is within the tube. If it is outside the tube, the loss is the magnitude of the difference between the predicted value and the radius ε of the tube. Both C and ε are user-determined parameters. Two positive slack variables ξ_i, ξ_i^* are used to cope with infeasible constraints of the optimization problem. To get the estimation of w and b , Equation (5) can be transformed to a primal objective function (7).

$$MinJ = \frac{1}{2} \|w\|^2 + C \sum_{i=1}^s (\xi_i^* + \xi_i) \tag{7}$$

$$s.t. \begin{cases} y_i - w \cdot \Phi(x_i) - b \leq \varepsilon + \xi_i^* \\ w \cdot \Phi(x_i) + b - y_i \leq \varepsilon + \xi_i \\ \xi_i^*, \xi_i \geq 0 \end{cases}$$

This constrained optimization problem is solved by using the following primal Lagrangian form:

$$L = \frac{1}{2} \|w\|^2 + C \sum_{i=1}^s (\xi_i^* + \xi_i) - \sum_{i=1}^s (\eta_i \xi_i + \eta_i^* \xi_i^*) - \sum_{i=1}^s \alpha_i (\varepsilon + \xi_i - y_i + w \cdot \Phi(x_i) + b) -$$

$$- \sum_{i=1}^s \alpha_i^* (\varepsilon + \xi_i^* - y_i - w \cdot \Phi(x_i) + b) \tag{8}$$

where L is the Lagrangian, and $\eta_i, \eta_i^*, \alpha_i, \alpha_i^*$ are Lagrange multipliers. Hence the dual variables in Equation (8) have to satisfy the positive constraints.

$$\eta_i, \eta_i^*, \alpha_i, \alpha_i^* \geq 0 \tag{9}$$

The above problem can be converted into a dual problem where the task is to optimize the Lagrangian multipliers, α_i and α_i^* . The dual problem contains a quadratic objective function of α_i and α_i^* with one linear constraint:

$$MaxJ = -\frac{1}{2} \sum_{i,j=1}^s (\alpha_i^* - \alpha_i)(\alpha_j^* - \alpha_j)(\Phi(x_i) \cdot \Phi(x_j)) + \sum_{i=1}^s \alpha_i^* (y_i - \varepsilon) - \sum_{i=1}^s \alpha_i (y_i + \varepsilon) \tag{10}$$

$$s.t. \begin{cases} \sum_{i=1}^s \alpha_i = \sum_{i=1}^s \alpha_i^* \\ 0 \leq \alpha_i \leq C \\ 0 \leq \alpha_i^* \leq C \end{cases}$$

$$Let \ w - \sum_{i=1}^s (\alpha_i - \alpha_i^*) x_i = 0 \tag{11}$$

$$Thus, \ f(x) = \sum_{i=1}^s (\alpha_i - \alpha_i^*) \Phi(x_i) \cdot \Phi(x_j) + b \tag{12}$$

By introducing kernel function $K(x_i, x_j)$, Equation (12) can be rewritten as follows:

$$f(x) = \sum_{i=1}^s (\alpha_i - \alpha_i^*) K(x_i, x_j) + b \tag{13}$$

where $K(x_i, x_j)$ is the so-called kernel function which is equal to the inner product of two vectors x_i and x_j in the feature space $\Phi(x_i)$ and $\Phi(x_j)$, that is $K(x_i, x_j) = \Phi(x_i) \cdot \Phi(x_j)$.

Without having to compute map $\Phi(x)$, the kernel function is proven to simplify the use of mapping. Some popular kernel functions are the linear kernel, polynomial kernel and the radial-basis function (RBF) kernel. Using different kernel functions, one can construct different learning machines with arbitrary types of decision surfaces. In general, the RBF kernel, as a non-linear kernel function, is a reasonable first choice [26, 27]. Thus, the RBF kernel is chosen in this work:

$$K(x_i, x_j) = \exp\left(-\frac{\|x_i - x_j\|^2}{2\sigma^2}\right) \tag{14}$$

where σ is a parameter which determines the area of influence this support vector has over the data space.

As user-determined parameters in the SVR model and the RBF kernel, C, ε and σ have to be defined before model estimation. To reduce the search space, referring to previous literature using SVM [28, 29], it is recommended to use the constraints of the three parameters which attribute respectively to the range $C \in [2^{-5}, 2^5], \varepsilon \in [2^{-13}, 2^{-1}],$ and $\sigma \in [0, 2]$. Then the

optimal values of these three parameters are calculated by employing exhaustive search method. The values with which the minimum RRMSE of the model is generated are taken as the optimal values. RRMSE (Relative Root Mean Square Error) is a statistic for examining the goodness-of-fit of a model. It is calculated by

$$RRMSE(\hat{\theta}) = \sqrt{\frac{1}{n} \sum_{j=1}^n \left(\frac{\hat{\theta}_j - \theta_j}{\theta_j} \right)^2}$$

5.2 Comparing SVR and Hazard model

Instead of applying the Hazard model [30], which is often used in duration modelling, this paper introduces SVR and compares the performances of these two methods in $TT_{\text{home-work}}$, $TT_{\text{work-home}}$ and $AD_{\text{before\&post}}$ model. By using `streg()` function in Stata and the SVM Toolbox for Matlab respectively, both the Hazard and SVR models are estimated [20, 31, 32]. The prediction accuracies of $TT_{\text{home-work}}$, $TT_{\text{work-home}}$ and $AD_{\text{before\&post}}$ models are shown in Table 4. As mentioned above, the lower value of RRMSE represents the higher goodness-of-fit of the model [33]. Since all the RRMSE values of the three SVR models are accordingly lower than that of the Hazard models, SVR is chosen to be employed in the travel time and activity duration modelling.

5.3 Estimation results

Beside $TT_{\text{home-work}}$, $TT_{\text{work-home}}$ and $AD_{\text{before\&post}}$ model, the estimation results of the other five travel time and activity duration prediction models are shown in Table 4. RRMSE indicates that the forecast accuracies of all the models are acceptable, with that of $AD_{\text{work-based}}$ model and $AD_{\text{before\&post}}$ model being lower than the others. The travel distance is an important factor in both travel time and activity duration decisions, as it has significant impact on all the dependent variables. Travel mode, a proxy of travel speed, has influence on all the dependent variables except travel time of work-based sub-trip as well as activity duration of before-work&post-work trips. One possible reason is that most of the work-based sub-trip and before-work&post-work trips are short-distance trips, and 55.40% of them are made by walking. This results in the insignificant effect of traffic mode on travel time and activity duration. Moreover, the results show that the departure time has impact on the travel time of home-to-work commute trip and work-based sub-tour, which indicates that the traffic condition at different times of day will affect the travel time. In addition, the influence of travel departure time (or stop arrival time) in activity duration models (including $AD_{\text{hwstop\&ADwhstop}}$ model and $AD_{\text{before\&post}}$ model) shows that commuters will decide how long to stay in an activity partly based on the arrival time.

6. PREDICTION OF THE COMMUTE ACTIVITY-TRAVEL AGENDA

The model system developed can be directly applied to predict daily commute activity-travel schedule.

Here is an example: the first member in the family with ID number 0101040**** in our sample, is a 35-year old blue-collar worker, who earns 1501-2500 RMB every month. He had a typical commute activity-travel pattern on the survey day, which is composed of one morning and evening commute trip with one stop respectively, two work-based sub-tours, a before-work tour and two post-work tours. To examine the developed model system, only the first work-based sub-tour and post-work tour will be considered. According to the basic travel information of this respondent (shown in Table 5), his real daily commute schedule can be mapped in Figure 5. In comparison, his predicted daily commute activity-travel schedule is reported in Table 5 and Figure 6.

The results indicate that the typical daily commute schedule can be predicted by the model system developed in this paper with high accuracy. The maximum error for travel time and activity duration prediction in the above example is 4.69 minutes and 13.74 minutes, while the minimum error is 0.27 and 0.03, respectively. Furthermore, the hit ratio of the travel departure time and stop arrival time choice models is 57%. This indicates that there are some limitations with the Ordered Probit model, which will be detailed later. As for two of the models that produced wrong predictions, i.e., $TD_{\text{work-home}}$ and $SA_{\text{work-home}}$, the forecasted time ranges are all later than the observed value of our sample data. One potential reason may be that this commuter did not choose a common schedule for work-to-home commute trip as well as the stop during this trip, i.e., these two time points are a little earlier than the regular times, which are commonly after 16:30 p.m. according to the survey data.

The results also reveal that the prediction accuracies of SVR models are higher than those of the Ordered Probit models. The main reason is that we divided the natural continuous departure and arrival time into several discrete time intervals, and the number of alternative time intervals is heuristically determined. More alternatives would potentially increase the prediction accuracy, but significantly increase the complexity of the model.

7. CONCLUSION

In this paper, a model system of commuters' daily activity-travel time allocation has been constructed. The Ordered Probit model is employed for travel departure time and stop arrival time forecast, and Support Vector Regression is used in travel time and activity

Table 4 - Estimation results of the travel time and activity duration prediction models

	Model	Statistics						Variable						
		Mean	Variance	Max	Min	RRMSE	Age	Gender	Income	Mode	Occupation	Trip purpose	Departure time ¹	Travel distance
Travel time prediction models	TT _{before}	Observed value	12.52	0.46	30.00	3.00	0.0586		✓	✓	✓	✓		✓
		Predicted value	12.37	0.20	25.00	2.90								
	TT _{home-work}	Observed value	20.64	1.97	70.00	5.00	0.1341	✓	✓	✓	✓	✓	✓	✓
		Predicted value	15.26	0.61	46.99	0								
	Hazard	Observed value	18.94	1.80	135.00	3.00	0.1424	✓	✓	✓	✓	✓	✓	✓
		Predicted value	19.38	3.30	201.10	7.77								
	TT _{work-based}	Observed value	15.74	2.39	90.00	2.00	0.1639	✓	✓	✓	✓	✓	✓	✓
		Predicted value	10.31	0.31	25.81	0								
	TT _{work-home}	Observed value	18.28	1.73	70.00	4.00	0.1421	✓	✓	✓	✓	✓	✓	✓
		Predicted value	14.17	0.50	27.00	0								
	Hazard	Observed value	18.82	1.87	150.00	3.00	0.1631	✓	✓	✓	✓	✓	✓	✓
		Predicted value	19.50	4.22	245.29	8.57								
	TT _{post}	Observed value	15.83	1.10	60.00	3.00	0.1020	✓	✓	✓	✓	✓	✓	✓
		Predicted value	12.45	0.27	25.96	0								
AD _{work-based}	Observed value	51.34	10.83	160.00	1.00	0.3893	✓	✓	✓	✓	✓	✓	✓	
	Predicted value	32.78	0.78	48.63	0.15									
Activity duration prediction models	AD _{hwhstop} &AD _{whstop} ²	Observed value	6.81	0.23	25.00	1.00	0.0518	✓		✓	✓	✓	✓	
		Predicted value	5.11	0.00	6.11	0.00								
	AD _{before} &AD _{post} ³	Observed value	42.30	9.22	145.00	5.00	0.3188	✓	✓	✓	✓	✓	✓	
		Predicted value	35.32	0.91	50.84	0.00								
Hazard	Observed value	44.15	11.07	190.00	1.00	0.3743	✓	✓	✓	✓	✓	✓		
	Predicted value	37.15	0.76	117.34	24.49									

¹ It refers to stop arrival time in the AD_{hwhstop}&AD_{whstop} model while representing departure time in other models;

² An extra category variable is introduced to this model besides variables in Table: 1 for stop duration in home-to-work commute trip and 2 for stop duration in work-to-home commute trip;

³ An extra category variable is employed in this model: 1 for before-work morning trip and 2 for post-work evening trip.

Table 5 - Observed and predicted values for daily commute time allocation

Activity-travel pattern	Basic travel information			Travel departure time and stop arrival time		Travel time		Activity duration	
	Travel distance (km)	Mode	Trip purpose	Observed value	Predicted value	Observed value (minute)	Predicted value (minute)	Observed value (minute)	Predicted value (minute)
Before-work trip	1.00	Walk	Maintenance	6:00	(6:00, 7:30]	20	17.00	20	19.97
Home-to-work commute trip	1.50	Walk	Work	7:35	(7:30, 9:00]	20	19.25	-	-
Home-to-work stop	0.30	- ¹	Maintenance	7:49	(8:00, 9:00]	-	-	6	7.55
Work-based sub-tour	0.50	Bus	Work	9:30	[9:00, 11:00]	5	9.69	45	31.26
Work-to-home commute trip	1.50	Auto	Other	15:47	(16:00, 17:30]	7	7.27	-	-
Work-to-home stop	0.16	-	Maintenance	16:02	(16:30, 18:00]	-	-	2	3.23
Post-work trip	0.20	Bus	Maintenance	18:30	[18:00, 19:00]	5	4.20	20	18.23

¹ "-" represents that there is no significant value for the item.

duration modelling. Time allocation of the typical daily commute activity-travel pattern is predicted with the model system.

This work develops a model system that can be used to predict the time allocation decisions of all the activity-travel segments (including stops) in a typical commute day, and it focuses on temporal pattern learning and modelling, which cannot only enhance the model's prediction accuracy, but also makes it more suitable for analysis of time-related policies. By introducing the Support Vector Regression, it also makes a methodological contribution in enhancing prediction accuracy of the travel time and activity duration prediction. This work also serves as a foundation on which future models of full-scale daily activity-travel pattern can be built.

Study results can be applied to a wide range of Transportation Demand Management (TDM) policies, especially the measures aimed at reducing commute trips, such as flexible working hours. Flexible working

hours provide commuters with a wide span of departure time choices. The model system can then be applied to predict the commuters' travel time based on each departure time alternative. The predicted travel time can be used to indicate the optimal departure time and thus the development of flexible working hour program targeting this departure time. This study is also essential for planning the development and construction of new transportation infrastructure by providing predicted temporal travel demands. Moreover, the temporal travel demand information can be integrated in the traveller information system to help people make better travel decisions.

One limitation of the current study is that the prediction accuracy of the discrete choice model is lower than the continuous model, as it divides the naturally continuous time into artificially defined time periods. To address this problem, we may try to use more continuous models instead of discrete choice models in daily time allocation. For example, we may employ SVR

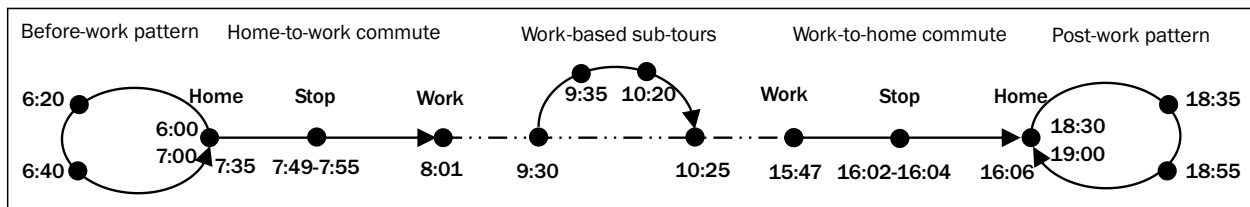


Figure 5 - Observed daily commute time allocation

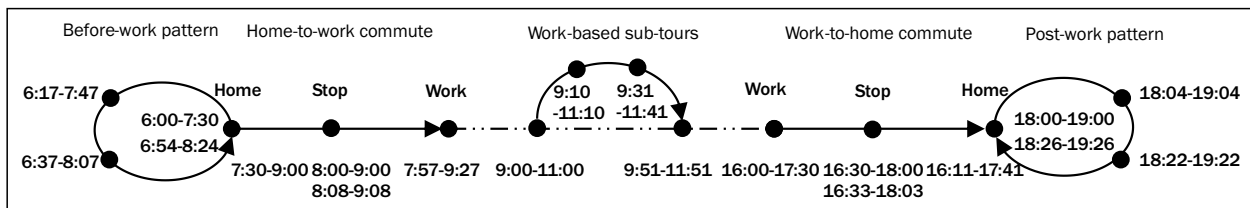


Figure 6 - Predicted daily commute time allocation

to examine how long it will take for a commuter to arrive to the first stop after leaving home for work. Then the stop arrival time of home-to-work trip can be calculated with travel departure time of home-to-work trip + this time interval, instead of being modelled with a discrete choice model directly. By changing the discrete models to continuous models, the prediction accuracy of the daily time allocation model system can increase.

ACKNOWLEDGEMENTS

The research is funded by the National Natural Science Foundation of China (50908099 and 51278221) and the Doctoral Program of Higher Education of China (201104493).

宗芳¹, 贾洪飞¹, 潘翔², 吴杨¹

¹吉林大学交通学院 长春 中国

²浙江工业大学计算机科学与技术学院 杭州 中国

日通勤时间预测

本文建立了通勤者日活动-出行时间预测模型系统。应用Ordered Probit模型预测了出发时刻和到达时刻,应用Support Vector Regression分析了出行时间和活动耗时,并将所建模型系统用于实际的出行时间预测实验,结果表明模型可以达到较高的预测精度。另外,通过考查各因素对通勤者出行时间决策的影响,本文还分析了通勤者的活动-出行时间决策机理。本文的研究可以通过提供可靠的出行需求时间预测结果,在解决实际交通问题中得到进一步应用,如建立出行信息系统等。另外,通过引入Support Vector Regression,本文还在理论层面提高了出行时间和活动耗时的预测精度。

关键词

时间分配; 通勤; 活动; 出行; Support Vector Regression

REFERENCES

- [1] Yu, B., Yang, Z.Z., Li S.: *Real-Time Partway Deadheading Strategy Based on Transit Service Reliability Assessment*, Transportation Research Part A, Vol. 46, No. 8, 2012, pp. 1265-1279
- [2] Zong, F., Juan, Z., Jia, H.: *Examination of staggered shifts impacts on travel behavior: a case study of Beijing*, Transport, Vol. 28, No. 2, 2013, pp. 175-185
- [3] Ettema, D., Bastin, F., Polak, J., Ashiru, O.: *Modelling the joint choice of activity timing and duration*, Transportation Research A, Vol. 41, 2007, pp. 827-841
- [4] Hamed, M.M., Mannering F.L.: *Modeling travelers' postwork activity involvement: toward a new methodology*, Transportation Science, Vol. 27, No. 4, 1993, pp. 381-394
- [5] Mao, B.H., Chen, J.C.: *2005 Beijing Household Travel Survey Report*. Beijing Transportation research Center, Beijing Jiaotong University, 2006, Beijing
- [6] Bowman, J.L., Ben-Akiva, M.E.: *Activity-based disaggregate travel demand model system with activity schedules*, Transportation Research Part A, Vol. 35, 2000, pp. 1-28
- [7] Bhat, C.: *Analysis of travel mode and departure time choice for urban shopping trips*, Transportation Research B, Vol. 32, 1998, pp. 361-371
- [8] Small, K.A.: *The scheduling of consumer activities: work trips*, American Economic Review, Vol. 72, 1982, pp. 467-479
- [9] Small, K.A.: *A discrete model for ordered alternatives*, Econometrica, Vol. 55, NO. 2, 1987, pp. 409-424
- [10] Bhat, C., Steed, J.: *A continuous-time model of departure time choice for urban shopping trips*, Transportation Research, Vol. 36B, 2002, pp. 207-224
- [11] Juan Z., Xianyu J.: *Daily Travel Time Analysis with Duration Model*, Journal of Transportation System Engineering & Intelligent, Vol. 10, No. 4, 2010, pp. 62-67
- [12] Pendyala, R., Bhat, C.R.: *An exploration of the relationship between timing and duration of maintenance activities*, Transportation, Vol. 31, 2004, pp. 429-456
- [13] Habib K.M.N.: *Modeling commuting mode choice jointly with work start time and work duration*, Transportation Research Part A, Vol. 46, 2012, pp. 33-47
- [14] Schwanen, T., Dijst, M.: *Travel-time ratios for visits to the workplace: the relationship between commuting time and work duration*, Transportation Research Part A, Vol. 36, 2002, pp. 573-592
- [15] Kitamura, R., Fujii, S.: *Two computational process models of activity-travel behavior*. In T. Garling, T. Laitila and K. Westin (eds.) *Theoretical Foundations of Travel Choice Modeling*, Oxford: Elsevier Science, 1998, pp. 251-279
- [16] Guo, J.Y., Bhat, C. R.: *Representation and analysis plan and data needs analysis for the activity-travel system*, Research Report, Texas department of transportation, 2001
- [17] Zong F., Lin H., Yu B., Pan X.: *Daily commute time prediction based on Genetic algorithm*. Mathematical Problems in Engineering, Vol. 2012, 2012, <http://dx.doi.org/10.1155/2012/321574>
- [18] Ray P.: *Independence of Irrelevant Alternatives*, Econometrica, Vol. 41, No. 5, 1973, pp. 987-991
- [19] Quddus, M.A., Noland, R.B., Chin, H.C.: *An analysis of motorcycle injury and vehicle damage severity using ordered probit models*, Journal of Safety Research, Vol. 33, No. 4, 2002, pp. 445-462
- [20] Lawrence C. Hamilton. *Statistics with Stata: Version 12*. Cengage, 2013
- [21] Este A., Gringoli F., Salgarelli L.: *Support Vector Machines for TCP traffic classification*, Computer Networks, Vol. 53, 2009, pp. 2476-2490
- [22] Anguita, D., Boni, A., Ridella, S.: *Evaluating the Generalization Ability of Support Vector Machines through the Bootstrap*, Neural Processing Letters, Vol. 11, 2000, pp. 51-58
- [23] Sung, K., Poggio, T.: *Example-based learning for view-based human face detection*. IEEE Transactions on Pattern Analysis and Machine Intelligence, Vol. 20, No. 1, 1998, pp. 39-50
- [24] Vapnik, V.: *The Nature of Statistical Learning Theory*, Springer Verlag, New York, 1995
- [25] Li, Y.M., Gong S.G., Liddell, H.M.: *Support Vector Regression and Classification Based Multi-view Face Detection and Recognition*, IEEE International Conference on Automatic Face and Gesture Recognition, 2000, pp. 300-305

- [26] **Dong, B., Cao, C., Lee, S.E.:** *Applying Support Vector Machines to Predict Building Energy Consumption in Tropical Region*, Energy and Buildings, Vol. 37, No. 5, 2005, pp. 545-553
- [27] **Yao, B.Z., Yang, C.Y., Yao, J.B., Sun, J.:** *Tunnel Surrounding Rock Displacement Prediction Using Support Vector Machine*, International Journal of Computational Intelligence Systems, Vol. 3, No. 6, 2010, pp. 843-852
- [28] **Yao, J.B., Yao, B.Z., Li, L., Jiang, Y.L.:** *Hybrid model for displacement prediction of tunnel surrounding rock*, Neural network world, Vol. 22, 2012, pp. 263-275
- [29] **Yu, B., Yang, Z.Z., Yao, B.Z.:** *Bus Arrival Time Prediction Using Support Vector Machines*, Journal of Intelligent Transportation Systems, Vol. 10, No. 4, 2006, pp. 151-158
- [30] **Anastasopoulos, P., Islam, M., Perperidou, D., Karlaftis, M.:** *Hazard-based analysis of travel distance in urban environments: longitudinal data approach*, Journal of Urban Planning and Development, Vol. 138, No. 1, 2012, pp. 53-61
- [31] *Statistical Computing Seminars Survival Analysis with Stata*. http://www.ats.ucla.edu/stat/stata/seminars/stata_survival/
- [32] *SVM-Support Vector Machines Software*. http://www.support-vector-machines.org/SVM_soft.html
- [33] **Yu, B., William, H.K.L., Mei, L.T.:** *Bus Arrival Time Prediction at Bus Stop with Multiple Routes*, Transportation Research Part C, Vol. 19, No. 6, 2011, pp. 1157-1170

