Drivers Route Switching Behavior Based on Group Classification

XUEQIN LONG, SHANSHAN LIU, HUAN ZHAO, AND MENG ZHOU

Key Laboratory of Transport Industry of Management Control and Cycle Repair Technology for Traffic Network Facilities in Ecological Security Barrier Area, College of Transportation Engineering, Chang’an University, Xi’an 710064, China

Corresponding author: Xueqin Long (xqlong@chd.edu.cn)

This work was supported in part by the National Key Research and Development Program under Grant 2019YFB1600500, in part by the Science Program of Shaanxi Province under Grant 2020JM-222, and in part by the Fundamental Research Funds for the Central Universities, Chang’an University (CHD), under Grant 300102210204.

ABSTRACT Drivers’ route switching behavior shows obvious difference when they face various traffic conditions. The article studies the drivers’ route switching behavior based on group classification. Questionnaire combining SP survey and RP survey is carried out to collect the drivers’ route choice behavior under the influence of individual attributes, daily travel characteristic and traffic conditions. Latent Class Model (LCM) is used to analyze the behavior characteristic. According to the goodness of models, drivers are divided into three categories. Drivers of sensitive pattern will switch route easily which is represented by young people with shorter driver-age. In contrast, drivers of unresponsive pattern will not switch routes easily and the pattern is represented by elder people with longer driver-age. Based on the drivers’ classification results, ordinal logistic model is established. According to the odds ratio of each variable, we find that, age, driver-age, nature of drivers, and travel frequency all affect drivers’ route switching behavior.

INDEX TERMS Travel behavior, route choice, group classification, latent class model, ordinal logistic model.

I. INTRODUCTION Currently, study of drivers’ route choice is always the hotspot issue of behavior research. Drivers will face various traffic conditions and may adjust their routes continuously. Therefore, on-trip switching behavior is just different with that of pre-trip. Primary factors that drivers are concerned about may be different [1].

Disaggregate model is always used to analyze the characteristic of travel behavior for traditional research [2], [3]. In actual traffic system, heterogeneity and homogeneity coexist in travel behavior. That is, in a certain group, drivers have the similar drive behavior. Meanwhile, drivers belonging to different groups behave divergently. The impact mechanism of the influential factors causes the existence of heterogeneity and homogeneity. The influence of a certain factor on decision behavior varies among groups. However, there are few references focusing on the influence’s diversity of different drivers.

In the article, drivers are classified as different groups, and route switching behavior of each group is analyzed combing the qualitative and quantitative methods. Then, the influence of all factors on different groups is further analyzed. We can extract some influence mechanism of factors on different drivers. Results can be used to help managers to release traffic information and formulate guidance strategies.

The remainder of the article is organized as follows. Several route choice behavior studies frameworks are introduced in Section 2. Section 3 explains the research method, and analyzes the classifying results using Latent Class Model (LCM). Based on the classifying results, section 4 establishes an ordinal logistic model to study the drivers’ route switching propensity. Section 5 discusses the results of drivers’ classification and logistic model. Section 6 draws the conclusions and significance of the article.

II. LITERATURE REVIEW

A. ROUTE SWITCHING BEHAVIOR STUDY Disaggregate model is usually applied to study the drivers’ behavior, including Logit model, Nested Logit model, Cross Nested Logit model, Mixed Multinomial Logit model, Latent Class Logit model [3]. Disaggregate model can quantitatively describe the decision results and analyze factors’ influence,
so it has been widely used in the area of behavior study. Considering the difference of perception ability and traffic conditions’ uncertainty, Bounded Ration (BR) is widely used instead of expected utility theory. Mahmassani studied the User Equilibrium under Bounded Rational theory first [4]. Then, Di and Liu [5], Ye and Yang [6] and Zhang et al. [7], [8] established boundedly rational user equilibrium to study the drivers’ route choice behavior. Evidences show that, bounded rational behavior exists in route choice.

Additionally, Regret Theory (RT) is also widely applied to drivers’ route choice behavior analysis. RT was proposed by Loomes in 1982 [9]. Li introduced regret aversion function and established a stochastic user equilibrium Logit model based on RT [10], [11]. Some research applying Game Theory study the route choice behavior [12].

For route choice analysis, methods of data collecting mainly include Stated Preference (SP)/ Revealed Preference (RP) survey, experiment, simulation, approaches based on data science and data-driven (for example, GPS trajectory data [13], vehicle test data [14], vehicle Telematics Data [15]). Comparing with traffic survey, experiment has stronger controllability. Behavior experiment can evaluate the effect of some non-implemented management policies, reveal the coupling mechanism of various factors through setting up control groups [16].

Some researchers analyze the factors influencing route switching behavior, such as the guidance information. SP survey and RP survey [17], [18] are both used to study the route switching behavior usually. Li found that, when drivers can accept the traffic information, they are more prone to change routes [19]. Route switching frequency is also depended on the information pattern, sensitivity on trip toll and drivers’ education level [20], and the attention paid to Variable Message Signs (VMS), distance between VMS and the destination, traffic condition of alternative route [21]. Additionally, driving style, travel time, credibility of VMS and accuracy of guidance information are major factors influencing drivers’ route switching propensity [22]. Ben adopted Logit model to analyze the influence of descriptive information, prescriptive information and post-choice feedback information on drivers’ route switching behavior, and results indicate that, prescriptive information is the critical factor [23].

Additionally, socio-economic attributes [24], accidents [25], driver’s familiarity with road network [26] and other drivers’ switching behavior [27] have influence on route choice also. Drivers who are unfamiliar with road network have more consistent behavior and usually show lower propensity to switch to other routes [26]. In addition, congestion state may affect drivers’ route choice [27].

### B. GROUP CLASSIFICATION

Group classification can be used to analyze the difference among different groups. Currently, it has been mainly applied on the classification of airport passengers and high-speed passengers. Focused on conventionally piloted aircraft (CPA) and remotely piloted aircraft (RPA), results reveal that model choice is sensitive to individuals’ attitude towards new technology [28]. Similarly, Qiao classified aviation passengers using Latent Class Model [29]. Based on the attributes such as age, gender, travel date, travel distance, way to get ticket and pre-purchase time, passengers of high-speed railway were classified as three categories [30]. Results show that, three categories of passengers have significant differences in terms of pre-purchase time and travel distance.

For commuters, drivers’ classification focuses on trip mode choice. Chen developed a SP survey considering economic level, commuting distance and cost, attendance, comfort, and classified commuters by travel mode choice [31]. Molin studied the multimodal travel groups based on the self-reported frequency of mode use, and estimated the membership function to predict the belonging probability [32]. Also, normative beliefs, modality styles both affect travel mode choice behavior [33].

According to the relative researches, regularity of drivers’ route choice has been revealed adequately. However, drivers’ heterogeneity and homogeneity of route switching are not fully displayed still. That is, the behavior’s difference among drivers should be further discussed, reason and commonness is needed be deeply revealed. Therefore, based on the references, the article tries to classify the drivers according to route switching behavior. Then, drivers’ behavior characteristic of each group under different multi-level factors will be analyzed.

### III. DRIVERS GROUP CLASSIFICATION

In the section, Latent Class Model is introduced to understand the method of group classification in the first part. Additionally, data collecting method are explained, attributes and implications are illustrated in detail. In the third part, Latent Class Model are established to estimate the parameters using the survey data. According to the calculating results, drivers are classified to the certain cluster. At the end of the section, characteristic of drivers’ route choice behavior is analyzed.

### A. LATENT CLASS MODEL

Latent Class Model (LCM) was earliest applied for psychology study by Lazarsfeld and Henry [34]. LCM can deal with different latent cluster variables, explain the inter-relationship between variables. LCM assumes that, any statistical result of the observation data can be categorized to a specific cluster. Each level of latent variables is completely independent, each observation data can only belong to a certain level. If $A$, $B$, $C$ are three manifest variables (that are questions in the questionnaire), the mathematical expression of LCM should include un-conditional probability representing potential categories ($\pi^A_{ijk}$) and conditional probability reflecting the latent cluster structural. The two expressions constitute the basic Latent Class Model.

$$
\pi^A_{ijk} = \sum_{t=1}^T \pi^x_{it} \pi^B_{it} \pi^C_{kt} \pi^x_{kt}
$$

(VOLUME 8, 2020) 93515
is the joint-probability, reflecting the probability of set \( \{i, j, k\} \) accounted for all observed data. Joint-probability is the sum of latent cluster probability.

\( \pi^t_{ijk} \) is the probability of observation data belonging to a specific latent cluster \( t \), that is \( P(X = t), t = 1, 2, \ldots, T \).

\( \pi^t_{it} \) is the conditional probability. It means the response probability of observation data belonging to the \( i \)th level of manifest variable \( A \) among cluster \( t \). That is \( P(A = i | X = t), X = 1, 2, \ldots, I \). And so on.

Maximum likelihood estimation is used to solve the LCM. For the LCM model with \( T \) latent clusters, maximum likelihood estimation function of latent variable \( X \) is followed as:

\[
\hat{\pi}^{ABC}_{ijkt} = \frac{\hat{\pi}^X_{it} \hat{\pi}^{TX}_{jt} \hat{\pi}^{RX}_{kt} \times \hat{\pi}^{CX}_{t}}{\hat{\pi}^{ABCX}_{itjkt}}
\]

(2)

\( \hat{\pi}^X_{it} \) is the probability estimation value. \( \hat{\pi}^{TX}_{jt}, \hat{\pi}^{RX}_{kt}, \hat{\pi}^{CX}_{t} \) are the latent cluster probability of manifest variables \( A, B, C \) respectively. The sum of probability estimation of each manifest variable for latent cluster \( T \) is maximum likelihood estimation (MLE) joint probability \((\hat{\pi}^{ABC}_{ijkt})\).

According to the characteristic of probability model, two constraint conditions should be satisfied.

\[
\sum^T_{t=1} \hat{\pi}^X_{it} = 1
\]

(4)

\[
\sum^T_{t=1} \hat{\pi}^X_{it} = \sum^T_{j} \hat{\pi}^{BX}_{jt} = \sum^T_{k} \hat{\pi}^{CX}_{t} = 1
\]

(5)

Maximum likelihood estimation (MLE) is used to solve the value of latent cluster probability \((\hat{\pi}^X_{it}, \hat{\pi}^{BX}_{jt}, \hat{\pi}^{CX}_{t})\). Maximum likelihood probability of manifest variables’ each level is obtained by dividing the above two formulations.

\[
\hat{\pi}^{ABCX}_{ijkt} = \frac{\hat{\pi}^{ABC}_{ijkt}}{\hat{\pi}^{ABC}_{ijkt}}
\]

(6)

When solving the model, assume the latent cluster is one. Then increase the clusters gradually, estimate parameters and make cluster analysis again.

AIC and BIC indicators are used to evaluate the goodness of fit of LCM model.

\[
AIC = 2K - 2 \ln(L)
\]

(7)

\[
BIC = K \ln(n) - 2 \ln(L)
\]

(8)

\( K \) is number of parameters in the model (indicating the complexity of the model). \( L \) is likelihood function. \( n \) is the number of samples. Less AIC and BIC mean better fitness. AIC prefers the proper model with less estimation parameters and larger degree of freedom (df). BIC prefers the proper model with less parameters. According to some researches, when samples exceed 1000, BIC can be used [35].

The final procedure of LCM is classifying all samples into the appropriate potential class, through calculating the posteriori probability of the samples belonging to class \( t \).

Assume a sample’ choice is \( i, j, k \) respectively for manifest variables \( A, B \) and \( C \), the probability \( \hat{\pi}^{ABCX}_{ijkt} (t = 1, 2, 3, \ldots T) \) denotes the joint probability of the samples belonging to latent class \( t \). Then we can calculate the posteriori probability of the sample belonging to class \( t \).

\[
\hat{\pi}^{ABCX}_{ijkt} = \frac{\hat{\pi}^{ABC}_{ijkt}}{\sum^T_{t=1} \hat{\pi}^{ABC}_{ijkt}}
\]

(9)

\( \hat{\pi}^{ABCX}_{ijkt} \) is the posteriori probability of the samples belonging to class \( t \). \( \hat{\pi}^{ABC}_{ijkt} \) is the joint probability of the samples choice, which can be obtained according to LCM (formulation 2).

B. DATA COLLECTING

Online questionnaire combing Stated Preference (SP) survey and Revealed Preference (RP) survey was used to collect the travel behavior data. The questionnaire includes three parts composed of 16 questions. All the questions are multiple-choice.

The first part are individual attributes. The second part is RP survey including daily travel attributes. Participants are required to answer the questions according to their daily travel behavior. The third part is SP survey involving the real-time traffic conditions that may influence route switch behavior and contains 7 scenarios. Respondents are required to answer, under which state, he/she will change his/her quorate? Attributes and implications are shown in Table 1.

The questionnaire is posted online (https://www.wenjuan.com), any participant can browser and answer the questionnaire. The questionnaire is placed at the top of the network through payment then more internet users will browser the questionnaire. The online recruit lasted for two weeks. 354 valid participants have been obtained.

Participants’ demographics are described in Table 2. Drivers less than 50 years old are accounted for 92.4%, and drive-age is most less than 3 years. Most participants (84%) are well educated.

The online investigation lasts for two weeks, and the sample distribution is shown as figure 1. Participants reaches to the peak at the forth day and continues to decrease from the seventh day.
Statements of participants’ daily travel attributes is shown as figure 2. Most participants travel less than two time within a day, and the travel time is no more than 2 hours.

We investigate the travelers’ route switching behavior under different traffic conditions. Figure 3 shows the influence of traffic conditions on participants’ route switching behavior.

We can find that, most travelers may change their route only when they are faced with serious accidents (fatal and extra serious accidents), and minor accidents will not induce the route switching behavior (figure 3(a)). Other travelers’ choice doesn’t really affect the route switching behavior (figure 3(b)). Figure 3(c) shows that, travelers are sensitive to traffic delay.

Cronbach’s alpha is used to evaluate the internal reliability of questionnaire, which refers to the average value of half confidence coefficient obtained by all possible division methods of the scale. Large Cronbach’s alpha means high reliability of the data. When the value of Cronbach’s alpha is distributed between 0.7 and 0.9, we can accept that the data is reliable. In the article, Cronbach’s alpha = 0.765 for the samples, therefore, the data is reliable.

Bartlett ball test and KMO (Kaiser-Meyer-Olkin) test are applied to evaluate the validity of questionnaire. When KMO is more than 0.8, factor analysis and samples classification...
FIGURE 2. Statistics about the participants’ daily travel attributes.

TABLE 3. KMO and Bartlett test.

| Test method     | Result         |
|-----------------|----------------|
| Kaiser-Meyer-Olkin KMO | 0.842 |
|                 | Chi square     | 3608.671 |
| Bartlett        | df             | 120      |
|                 | Sig.           | 0.000    |

are compatible for the data. Significant level of Bartlett is less than 0.05, we can assume the data is effective. Analysis result is showed in Table 3, which indicates the data is valid.

C. GROUP CLASSIFICATION BASED ON LATENT CLASS MODEL

The manifest attributes of the model include sixteen variables shown as Table 1.

1) GOODNESS OF FIT OF THE MODEL

In order to select the most appropriate LCM, number of latent categories is gradually increased from 1. Five LCM models are established and goodness of fit is tested respectively, the results are shown in Table 3. We can find that, Log Likelihood(LL) declines with the increases of cluster number. At the same time, AIC declines and BIC ascends. Model 3 has relative lower AIC and BIC. P-value of model 3 is less than 0.05, and the model is significant. Then samples are categorized as three-cluster, and model 3 is chosen for further analysis.

2) ESTIMATION RESULTS

Explanatory strength of latent cluster for every variable is listed in Table 4. Manifest variables with poor explanatory(P>0.05) have to be excluded. Factors are analyzed after deleting the four variables, results are shown in Table 5.
TABLE 4. Goodness of fit.

| Model | Cluster | LL    | BIC(LL) | AIC(LL) | L/df  | P-value |
|-------|---------|-------|---------|---------|-------|---------|
| Model 1 | 1-Cluster | -1769.08 | 3738.5351 | 3622.166 | 2412.2848 | 1.2e-453 |
| Model 2 | 2-Cluster | -1728.56 | 3738.5919 | 3575.1215 | 2331.2399 | 9.4e-450 |
| Model 3 | 3-Cluster | -1694.32 | 3751.2279 | 3540.6559 | 2262.7743 | 2.2e-449 |
| Model 4 | 4-Cluster | -1660.44 | 3764.5544 | 3506.8808 | 2194.9992 | 1.6e-450 |
| Model 5 | 5-Cluster | -1639.31 | 3803.4013 | 3498.626 | 2152.7444 | 7.2e-460 |

TABLE 5. Estimation-model 3.

| Variables | Cluster 1 | Cluster 2 | Cluster 3 | Wald  | p-value | R²  |
|-----------|-----------|-----------|-----------|-------|---------|-----|
| age       | 0.7301    | -2.08     | 1.3499    | 19.4789 | 5.90e-05 | 0.3397 |
| d-age     | 0.2267    | -0.973    | 0.7463    | 14.0925 | 0.00087 | 0.281 |
| job       | 0.0268    | 1.3574    | -1.3942   | 21.7629 | 1.90e-05 | 0.3712 |
| nature    | -0.5215   | -0.2585   | 0.78      | 51.1   | 0.0078 | 0.1556 |
| num       | -0.6355   | 0.5882    | 0.0473    | 68.02  | 0.00033 | 0.2779 |
| accident  | -0.0908   | 0.0947    | -0.0038   | 23.68  | 0.00021 | 0.3431 |
| delay     | -2.2617   | 0.5174    | 1.7444    | 15.915 | 0.00034 | 0.4461 |
| other-d   | -0.0836   | 0.5888    | -0.4852   | 16.378 | 0.0057 | 0.2601 |
| percent   | -0.5453   | -0.241    | 0.7863    | 11.9809 | 0.0025 | 0.188 |
| con       | -0.9119   | 0.1742    | 0.7377    | 25.6185 | 0.0018 | 0.1747 |
| con-con   | -0.906    | -0.0143   | 0.9204    | 15.9894 | 0.00034 | 0.2363 |
| familiar  | 0.1076    | -0.3712   | 0.2636    | 12.635 | 0.0026 | 0.1283 |

3) LATENT CLUSTER PROBABILITY

Table 6 shows the latent cluster probability of LCM. Cluster size is latent cluster probability, other values in the table denote the condition probability. The probabilities of drivers belonging to three clusters are 48.49%, 41.25% and 10.26% respectively. The first two clusters are dominant.

We can conclude from the latent cluster probability, the group of cluster 2 is very sensitive to traffic conditions and is prone to switch their routes frequently. The group focuses on youngsters with shorter drive-age and has unsteady job. In contrast, drivers of cluster 3 have very low switching propensity, they will not switch the current routes until the road conditions are very terrible. Propensity to change route for cluster 1 is in the middle of cluster 2 and 3.

Based on the analysis above, three groups are named as steady group for cluster 1, sensitive group for cluster 2, unresponsive group for cluster 3. Three groups are accounted for 48.49%, 41.25% and 10.26% respectively. Steady and sensitive groups have similar proportion respectively.

4) CLASSIFICATION ACCURACY

All samples’ posteriori probabilities can be calculated and shown as Table 7.

Take the first sample as example, the posteriori probability of three classes is 0.2851, 0.7149 and 0 respectively. The posteriori probability of cluster 2 is obvious larger than other clusters. Therefore, we can classify the first sample 1 into cluster 2. The actual categories of all samples are shown in Table 8.

There is little difference between the actual category probability and latent class probability. Then, the category results have high accuracy and the category results are reasonable.

D. BEHAVIOR CHARACTERISTICS ANALYSIS

Further, analyze the travel characteristics of three clusters. Focus on the third part of the questionnaire, analyze the factors affecting drivers’ route switching. Results are shown as figure 4.

General accidents have greatest impact on drivers’ behavior and will lead about 52% drivers to switch routes. Slight accidents have great impact on drivers belonging to cluster 2 but less impact on cluster 3. Drivers belonging to cluster 3 are more sensitive to serious accidents.

For delay time, slight delay will lead about 50% drivers belonging to cluster 2 to switch routes. All drivers belonging to cluster 2 will change routes when delay is about 40 minutes. Drivers of cluster 3 have little reaction on slight delay. Sensitivity of cluster 1 is in-between.

Drivers of cluster 2 are sensitive to other’s decisions. When less than 20% other drivers change routes, about 50% drivers belonging to cluster 2 will change routes. However, when more than 60% other drivers change routes, about 40% drivers belonging to cluster 3 will change routes.

Influence of congestion frequency has the similar tendency as delay time.

No matter which clusters the drivers belonging to, drivers may change route only when they are familiar with the network. Which means, both sensitive and unresponsive groups are highly dependent on their existing knowledge and experience.

According the analysis above, drivers belonging to cluster 2 (sensitive pattern) are very sensitive to the change of traffic conditions. We can conclude that, sensitive drivers have a lower acceptable threshold for travel delay. In travel process, they switch routes frequently to achieve their desired goals of minimizing the travel time. Sensitive drivers are adventurous. In contrast, drivers of cluster 3 (unresponsive pattern) have greater tolerance of travel time, they don’t switch route easily. Unresponsive drivers are conservative. Cluster 1 (steady pattern) is in-between of the other two clusters. They need to judge the traffic information to make the decision. They will change routes only when they think congestion is serious. Cluster 1 and cluster 2 are accounted for higher probability, which indicates most drivers can’t endure serious congestion.

IV. MODELING FOR THE ROUTE SWITCHING BEHAVIOR

In the section, ordinal logistic model is introduced in the first part. The category results of all samples obtained
by table 6 are considered as dependent variable and incorporated into ordinal logistic model. Based on which, drivers route switching model is established. In the second part, influencing factors are quantitatively analyzed after

\[
\ln \left( \frac{p(y = \text{low probability of route change})}{1 - p(y = \text{low probability of route change})} \right)
\]

\[
= -0.981 + 1.421 \times (\text{age} = 1) + 1.111 \times (\text{age} = 2)
\]

\[
+ 2.903 \times (\text{num} = 1) + 2.423 \times (\text{num} = 2) + 1.959 \times (\text{time} = 1) + 1.261 \times (\text{time} = 2)
\]

\[
- 2.241 \times (p - \text{time} = 1) - 2.157 \times (p - \text{time} = 2)
\]

\[
\ln \left( \frac{p(y = \text{general probability of route change})}{p(y = \text{low probability of route change})} \right)
\]

\[
= 2.524 + 1.421 \times (\text{age} = 1) + 1.111 \times (\text{age} = 2) + 2.903 \times (\text{num} = 1) + 2.423 \times (\text{num} = 2)
\]

\[
+ 1.959 \times (\text{time} = 1) + 1.261 \times (\text{time} = 2) - 2.241 \times (p - \text{time} = 1) - 2.157 \times (p - \text{time} = 2)
\]
A. ORDINAL LOGISTIC MODEL

Assume \( y \) is ordinal dependent variable with \( n \) levels, \( X = x_1, x_2, \ldots, x_m \). Independent variables can be classified or continuous. The probability of \( y \) belonging to level \( j \) \((j = 1, 2, \ldots, n)\) is \( p(y = j|x) \). \( n \) levels can be divided as two classes: \{1, 2, \ldots, \( j \)\} and \{(\( j + 1 \)), (\( j + 2 \)), \ldots, \( n \)\}, \( j = 1, 2, \ldots, n - 1 \). Then, \( n - 1 \) two-class logistic regression models should be fitted.

\[
\ln \left( \frac{\sum_{i=1}^{j} p_i}{1 - \sum_{i=1}^{j} p_i} \right) = \alpha_j + \sum_{i=1}^{m} \beta_i x_i \quad (10)
\]

\( p \) is cumulative probability, \( \alpha_j \) is estimation constant, and \( \beta_i \) is regression parameter. \( \exp(\beta_i) \) is odds ratio (OR), which denotes the multiples of dependent variable \( y \) increases one level, when the independent variable \( x_i \) increases one unit. If variable \( x_i \) has \( r \) levels, the regression model should include \( r - 1 \) independent variables.

In the article, \( y \) means drivers’ group. We need to transfer the category result of each sample obtained by LCM in Table 6 to \( y \) that is needed in the Ordinal Logistic Model. Transfer regulation: Category result \( = 3 \) denotes drivers with lowest route switching frequency, corresponding to \( y = 1 \). Category result \( = 1 \) denotes drivers with general route switching frequency, corresponding to \( y = 2 \). Category result \( = 2 \) denotes drivers with highest route switching frequency, corresponding to \( y = 3 \).

The individual attributes data of first part, daily travel attributes of second part in the questionnaire and new added dependent variables \( y \) are used to set up ordinal logistic model. Then analyze the influence of individual attributes and daily travel attributes on route switch behavior according to parameters estimation.

B. RESULT ANALYSIS

Successive regression is applied, some variables with no statistical significance are eliminated step by step. Some variables with greater \( p \) (\( p > 0.25 \)) are eliminated and then remodel the reserved variables. Finally, variables of gender, driver-age, job, education level are excluded, and daily travel attributes are all maintained in the model. Parameters estimation of remodeling are shown as Table 9.

We can conclude \( \ln \left( \frac{1}{1 - p(y=low\ probability\ of\ route\ change)} \right) \), as shown at the bottom of the previous page.
Further analyze the OR (that is $\text{Exp}(\beta)$) of each variable. OR of variable $\text{age} = 1$ (corresponding to drivers less than 30 years old) is 4.142 (OR > 1). Which means, for drivers less than 30 years old, the possibility of route switching frequency increase is 4.142 times as drivers of 50 years old (OR = 1 for variable $\text{age} = 4$). Variable $\text{age} = 2$ has the similar trend. OR of $\text{age} = 3$ (corresponding to drivers of 41-50 years old) is 0.529, which means drivers of 41-50 years old have lower frequency of switching routes compared with drivers over 50 years old.

ORs of variable $\text{nature} = 1$ and $\text{nature} = 2$ are both less than 1, reveals that irritable drivers ($\text{nature} = 3$) have high frequency of switching routes.

ORs of variable $\text{num}$ are all more than 1, which means probability of route switching will reduces when travel frequency increase. For drivers with fewer travel times,
The possibility of route switching frequency increase is more than 10 times (OR = 18.222 for num = 1 and OR = 11.274 for num = 2) as drivers traveling once a day(num = 3). Feature of variable time reflects the similar feature as variable num. Longer travel time will lead to lower frequency of switching route.

ORs of variable p-time are less than 1, which means longer travel time in peak hour will lead to higher probability of switching route.

According to the results, drivers less than 40 years old will switch routes more frequently than drivers over 50 years old. However, drivers of 41-50 years old show lowest probability of route switching. That is to say, drivers of 41-50 years old have the strongest acceptance ability of travel time, then drivers over 50 years old are following. Threshold of travel time for young drivers is small so they prefer to change their routes frequently to look for the shortest route. For drivers’ character, irritable drivers have the highest frequency to switch routes(OR = 1), calm drivers will not change routes easily(OR = 0.134). For daily travel information, lower travel frequency and longer travel time will both lead to higher probability to change routes.

### C. MODEL TEST

Likelihood log ratio and goodness of fit test are showed as Table 10 and 11 respectively. Significance of likelihood log ratio test is 0, then the model is statistically significant.
Significance of pearson and deviance test are both 0, the model has high fitness.

V. DISCUSSION

(1) Combing results of LCM and ordinal logistic model, analyze the factors influencing on drivers’ route switching behavior.

① Individual attributes. Results of LCM indicates that, drivers of cluster 2 (sensitive group) are mainly young and irritate people and they have obvious tendency to switch route. The results are generally coherent with ordinal logistic model. Further, ordinal logistic model thinks there exists non-liner relationship between age and route change. That suggests, probability of route switching is not monotonous declined when drivers’ age increase. Drivers of 40-50 years old have the lowest probability to change route. That is the difference with results of LCM. Additionally, results of LCM indicate that, drivers with less driver-age and unsteady occupation are very sensitive. However, the two factors have no relationship with route switching behavior for logistic model. Two models both think education level and gender have no impact on route switching behavior. The results are different from Ma et al. [20], who find that the education level is the critical influence.

② Daily travel attributes. Results of LCM indicate that, only travel frequency in a week has influence on drivers’ route switching. Drivers with lower travel frequency prefer to switch routes frequently. The conclusions are consistent with ordinal logistic model. Moreover, travel frequency, average travel time and travel time in peak-hour all affect drivers’ route switching behavior in ordinal logistic model. The possibility to switch route for drivers with lower travel frequency is 10 times as the drivers travelling every day. Longer travel time and shorter travel time in peak-hour will both lead to lower frequency of route switching. Result of travel time analysis is consistent with Wang et al. [22], both confirm the influence of travel time. However, for drivers with long travel time, the prone to change route in the article (OR is more than 3) is more obvious than the results of Wang (OR is between 2 and 4).

③ Traffic environment. Analysis is only aimed at LCM. Factors such as accidents, delay, other drivers’ behavior, familiarity all affect drivers’ route switch behavior. No matter sensitive or unresponsive drivers, they will switch routes only when they are familiar with the network. The results are consistent with previous studies [28]–[30].

(2) Mechanism of influential factors.

① Individual attributes. To a certain extent, there is relationship with age and nature of drivers. Young drivers are more adventurous. Then they are prone to adjust the routes constantly to achieve the target of minimum travel time.

② Daily travel attributes. Frequent travel indicates the drivers are more familiar with the traffic conditions. Through continuous learning, drivers will choose the shortest route before departure. That is, the current route must be the shortest at common conditions. Then, even the traffic congestion happens, drivers will not switch routes easily. Travel time and peak-hour travel time show opposite trends, which means drivers all have a tolerance threshold of travel time. At peak hour, traffic congestion will lead to extremely long travel time. The increased travel time exceeds the drivers’ tolerance threshold, drivers may change routes to reduce travel time.

For peak-hour travel time, there is little research in current study. The results are opposite to common travel time. That is a new finding, which indicates, drivers have different behaviors among the different periods in a day. Overall, the behavior is related to traffic conditions. However, according to resident trip survey, the peak-hour travel time will not exceed one hour generally in China. Then, there is no significance of ORs for \( \text{num} = 2 \) and \( \text{num} = 3 \). That is the defect of questionnaire design and need to improve further. Therefore, the conclusion can be merged with travel time.

③ Traffic conditions. With the change of traffic conditions, sensitive drivers show obvious behavioral change. The unresponsive drivers show invariable behavioral characteristic.

The conclusions may be different from other references. We can find that, for drivers belonging to different groups, the trend and extent of impact of a certain factor are different. In the traditional research, we only analyze the influence of all drivers instead of the groups, then the single trend and extent of impact were obtained. We can conclude that, drivers with different ages and character behave diversely when they face traffic congestion. And the same conclusions can be obtained for drivers with different travel frequency and travel time. Findings of the article may conform to the facts accurately.

(3) Comparison with different latent class models.

① Estimation results of three models are shown as Table 12. For model 2, attribute nature, num, percent and familiar are not significant and deleted from the model. \( R^2 \) of the other attributes are less than that of model 3, which shows that, most attributes cannot explain the model better. For model 4, \( R^2 \) of the other attributes are larger than that of

| Results                        | Model 3 | Model 2 | Model 4 |
|-------------------------------|---------|---------|---------|
|                               | Cluster 1 | Cluster 2 | Cluster 3 | Cluster 1 | Cluster 2 | Cluster 1 | Cluster 2 | Cluster 3 | Cluster 4 |
| Actual category probability   | 53.4%    | 37.3%   | 9.3%    | 60.4%    | 39.6%    | 40.1%    | 36.3%    | 12.5%    | 11.1%    |
| Latent class probability      | 48.5%    | 41.2%   | 10.3%   | 53.9%    | 46.1%    | 49.6%    | 39.8%    | 6.2%     | 4.4%     |

TABLE 13. Category probability for each cluster for model 2 and model 4.
model 2 but less than model 3. The comparison results prove that model 3 is more appropriate than others, and the drivers should be classified as 3 clusters.

2) Accuracy of different latent class models.

Accuracy of three models are summarized as Table 13.

The classification error of model 2 is similar with model 3, but model 4 shows larger classification error. We can infer that; more categories will induce larger classification error. Then, 3-cluster is best for the survey results.

VI. CONCLUSIONS

Route switching behavior of drivers belonging to different groups is studied in the article. Drivers are classified as three clusters according to the probability of route switching. Drivers of different clusters behave diversely for a certain factor, and the results differ from the current study. In current study, all drivers are always considered as a whole and the overall behavior is analyzed. Now, we can include that, drivers show homogeneity among a certain cluster and heterogeneity between different clusters. Both the results of LCM and logistic model show that drivers’ individual attributes, daily travel information and traffic condition all affect drivers’ route switching behavior. Some new findings are shown as follow.

(1) The influence of age shows three obvious phases and there exists two inflection points. Probability of route switching is not monotonous declined when drivers’ age increase. The new finding is not involved in other references. Results in the article show that, route switching behavior displays different characteristic for drivers with different ages.

(2) Travel experience displays different impacts on drivers’ route switching behavior. If drivers have frequent daily travel, they are prone to keep the original routes unchanged. Which indicates drivers with rich travel experience always have relatively constant travel route. Travel frequency has not been specifically quantitative analyzed in current study. According to the article, route switching probability of drivers with lower travel frequency is more than 10 times than the drivers who travel every day(workday).

(3) For the ordinal logistic model, it can quantitatively describe the impact of all factors. But LCM can only provide a qualitative analysis. Then the combination of LCM and ordinal logistic model can effectively analyze the behavior characteristic. Some studies collect data through asking the route switching frequency of respondents. The frequency can’t denote the route change behavior during a long time. So, calibration the route switching frequency through some other observable factors by LCM can be more appropriate.

According to the findings, managers should provide differential traffic strategies at different period. At the peak-hour, most drivers are commuters with rich experience. They have relatively constant travel route and will not change the original route easily. Then, the traffic information will not impact the drivers’ travel behavior effectively. For managers, then should pay more attention on traffic guidance and let the traffic flow smoothly as soon as possible. Therefore, managers should focus on some key points, such as intersections, entrance and exit of the expressways and car park, bottleneck of the road section. At non-peak-hour, managers can provide some guidance information in variable massage sign. Then, traffic information is the primary factor to alleviate congestion.

ACKNOWLEDGMENT

The authors would like to thank all participants for their time and all colleagues of Traffic Engineering Department, School of Highway, Chang’an University.
[30] K. Qiao, P. Zhao, and J. X. Wen, “Passenger market segmentation of high-speed railway based on latent class model,” J. Transp. Sys., vol. 17, no. 1, pp. 29–34, 2017.

[31] J. B. Chen, “A study on heterogeneity of commuters’ choice of transport based on latent class models,” M.S. thesis, Transp. Eng., Beijing Jiaotong Univ., Beijing, China, 2016.

[32] E. Molin, P. Mokhtarian, and M. Kroesen, “Multimodal travel groups and attitudes: A latent class cluster analysis of dutch travelers,” Transp. Res. A, Policy Pract., vol. 83, pp. 14–29, Jan. 2016.

[33] R. Krueger, A. Vij, and T. H. Rashidi, “Normative beliefs and modality styles: A latent class and latent variable model of travel behaviour,” Transportation, vol. 45, no. 3, pp. 789–825, 2018.

[34] P. F. Lazarsfeld and N. W. Henry, Latent Structure Analysis, M. Houghton and J. A. Nelder, Eds. Boston, MA, USA: Houghton Mifflin, 1968.

[35] T. H. Lin and C. M. Dayton, “Model selection information criteria for non-nested latent class models,” J. Educ. Behav. Statist., vol. 22, no. 3, pp. 249–264, 1997.

---

XUEQIN LONG received the Ph.D. degree from the Beijing University of Technology (BJUT), in 2013. She is currently an Assistant Professor with the College of Traffic Engineering, Chang’an University. Her main research interests include travel behavior and travel psychology.

SHANSHAN LIU received the B.Eng. degree in aircraft propulsion engineering from Chang’an University, Xi’an, China, in 2017, where she is currently pursuing the M.A. degree in transportation planning and management. Her current research interests include travel behavior, autonomous driving, and AI algorithm.

HUAN ZHAO received the B.Eng. degree in aircraft propulsion engineering from Chang’an University, Xi’an, China, in 2018, where she is currently pursuing the M.A. degree in transportation planning and management. Her current research interests include travel behavior and traffic safety.

MENG ZHOU received the B.Eng. degree in aircraft propulsion engineering from Chang’an University, Xi’an, China, in 2017, where she is currently pursuing the M.A. degree in transportation planning and management. Her current research interests include travel behavior and public transport.

---

[20] F. Ma, Q. Q. Wu, and L. Wang, “Impact of multi-source guidance information on path change of drivers,” J. Transp. Syst., vol. 14, pp. 61–66, 2014.

[21] X. Chen, “Research on the new road network variable message sign’s impact on route choice,” M.S. thesis, Nat. ITS Center Eng. Technol. Res. Inst. Highway M.O.C., Beijing, China, 2008.

[22] X. Chen, “Research on the new road network variable message sign’s impact on route choice,” M.S. thesis, Nat. ITS Center Eng. Technol. Res. Inst. Highway M.O.C., Beijing, China, 2008.

[23] F. Ma, Q. Q. Wu, and L. Wang, “Impact of multi-source guidance information on path change of drivers,” J. Transp. Syst., vol. 14, pp. 61–66, 2014.

[24] X. Chen, “Research on the new road network variable message sign’s impact on route choice,” M.S. thesis, Nat. ITS Center Eng. Technol. Res. Inst. Highway M.O.C., Beijing, China, 2008.