

XIAN LI, M.Sc.^{1,2}

E-mail: 15120846@bjtu.edu.cn

HAIYING LI, Ph.D.¹

E-mail: hyl@bjtu.edu.cn

XINYUE XU, Ph.D.¹

(Corresponding author)

E-mail: xxy@bjtu.edu.cn

¹ State Key Laboratory of Rail Traffic Control and Safety

Beijing Jiaotong University

Beijing 100044, China

² School of Traffic and Transportation

Beijing Jiaotong University

Beijing 100044, China

Traffic Policy

Preliminary Communication

Submitted: 11 July 2017

Accepted: 4 Apr. 2018

A BAYESIAN NETWORK MODELING FOR DEPARTURE TIME CHOICE: A CASE STUDY OF BEIJING SUBWAY

ABSTRACT

Departure time choice is critical for subway passengers to avoid congestion during morning peak hours. In this study, we propose a Bayesian network (BN) model to capture departure time choice based on data learning. Factors such as travel time saving, crowding, subway fare, and departure time change are considered in this model. K2 algorithm is then employed to learn the BN structure, and maximum likelihood estimation (MLE) is adopted to estimate model parameters, according to the data obtained by a stated preference (SP) survey. A real-world case study of Beijing subway is illustrated, which proves that the proposed model has higher prediction accuracy than typical discrete choice models. Another key finding indicates that subway fare discount higher than 20% will motivate some passengers to depart 15 to 20 minutes earlier and release the pressure of crowding during morning peak hours.

KEY WORDS

departure time choice; Bayesian network; congestion; subway passengers;

1. INTRODUCTION

Passenger congestion causes a safety hazard during morning peak hours in subway stations. Capacity improvement is indeed an effective way to solve the problem; however, it is an impossible task at present because of its long construction period, high cost, and physical restrictions of stations [1]. A new method of analyzing departure time choice behavior and adjusting the attributes such as travel time and travel cost to guide travelers to change their departure time is put forward, aiming to decrease the total volume of travel demand during peak hours [2-3].

There are three conventional theories to model travel choice behaviors: expected utility theory, prospect theory, and regret theory [2-10]. For departure time choice, the expected utility theory has been

successfully adopted together with a discrete choice model, including multinomial logit (MNL) [3-4], nested logit (NL) [5], mixed logit (ML) [6-7], and probit models [8]. Bajwa [5] studied the departure time choice of car and rail commuters using NL, cross-nested logit, and ML models. Thorhauge et al. [3] analyzed the departure time choice of drivers and public transport commuters using the structural equation model and MNL, which suggested that fixed start time of work had a strong effect on departure time choice. Unfortunately, these models cannot address the nonlinear feature of departure time choice in subways in a quality manner.

The Bayesian network (BN), different from the discrete choice model with linear constraints, is a new approach to model travel choice behavior [11-14]. One of its major advantages is the intuitive and graphical representation of the causal relationships between data, which allows for better understanding [12]. Zhu et al. [11] proposed a mixed BN to model departure time and mode choice behavior of car users, which performed better than the NL model. Nozick et al. [13] developed a BN model for the travel mode choice problem, which identified key factors that influenced travel decision behaviors, such as underlying socioeconomic attributes and level of service. Nonetheless, there are not yet enough studies focusing on departure time choice by BN.

The aim of this paper is to develop a BN method to model departure time choice of subway passengers. We first conducted a D-optimal design SP survey and collected enough valid data to estimate the proposed BN model. Then, the K2 algorithm was employed to learn the model structure, and MLE was adopted to obtain its parameters from this data. Further, two different structures of BN are proposed to verify our method from the point of prediction accuracy, compared with discrete choice models.

This paper is organized as follows: Section 2 presents a brief introduction of BN; Section 3 provides a new BN approach that models departure time choice of subway passengers, where algorithms to determine model structure and parameters are also developed. A discussion on preferences of subway passengers under different attribute levels is provided in Section 4. Conclusions are drawn in Section 5, as well as future research directions.

2. DEPARTURE TIME CHOICE DATA COLLECTION

2.1 SP survey design





To collect the departure time choice data, we designed a SP survey. A passenger flow survey of Beijing Subway was first conducted to identify the factors of departure time choice and their relations from 6:30 am to 9:00 am during April to June 2015, involving more than 30 subway stations where inbound passenger

volume is high and passenger flow control measures are taken during peak hours. According to the survey result as shown in *Table 1*, we have:

- 1) If passengers choose usual departure time during peak hours, they will face serious crowding and long travel time.
- 2) If they choose earlier or later departure times, they may enjoy some travel time saving, as well as less crowding while having to afford another time cost.

Several trip-related attributes were investigated from a sample of 180 passengers using the Likert scale. Then, subway fare, crowding, travel time saving, and departure time change were selected as attributes of the stated choice (SC) experiment using the TOPSIS (technique for order preference by similarity to ideal solutions) method, and SC experiment was optimized using the D-optimal method [15]. Twelve choice sets were then generated, and one is shown in *Figure 1*, where respondents were required to choose their preferred departure time according to scenarios of subway fare, travel time saving, and crowding.

Table 1 - Results of passenger flow survey in Beijing Subway

| Period | Travel time cost | Crowding (in the most serious scenario) | Crowding (in the least serious scenario) |
|---|---|---|--|
| Peak hours (lasting for 30~50 minutes) | It will cost extra 5~10 minutes because of passenger flow control and a large number of passengers. | More than 20 passengers in a queue in front of a train door.  | 6~10 passengers in a queue in front of a train door.  |
| Off-peak hours | It will save 5~10 minutes because of decrease passengers flow and save travel time about 10%~20%. | 6~10 passengers in a queue in front of a train door.  | Less than 5 passengers in a queue in front of a train door.  |

Note: Average travel time is 52 minutes according to the statistics of Beijing Subway.



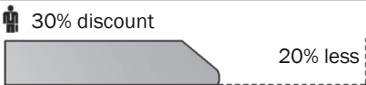

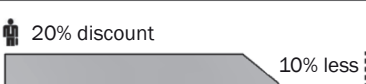

| Choice <input checked="" type="checkbox"/> | Subway fare | Travel time saving | Crowding |
|---|--|--------------------|---|
| Usual departure time <input type="checkbox"/> |  Full fare | | 10~20 passengers in a queue at train doors.  |
| Departure 20 min earlier than usual <input type="checkbox"/> |  30% discount | 20% less | 6~10 passengers in a queue at train doors.  |
| Departure 20 min later than usual <input type="checkbox"/> |  20% discount | 10% less | No more than 5 passengers in a queue at train doors.  |

Figure 1 – A scenario choice set in the survey questionnaire

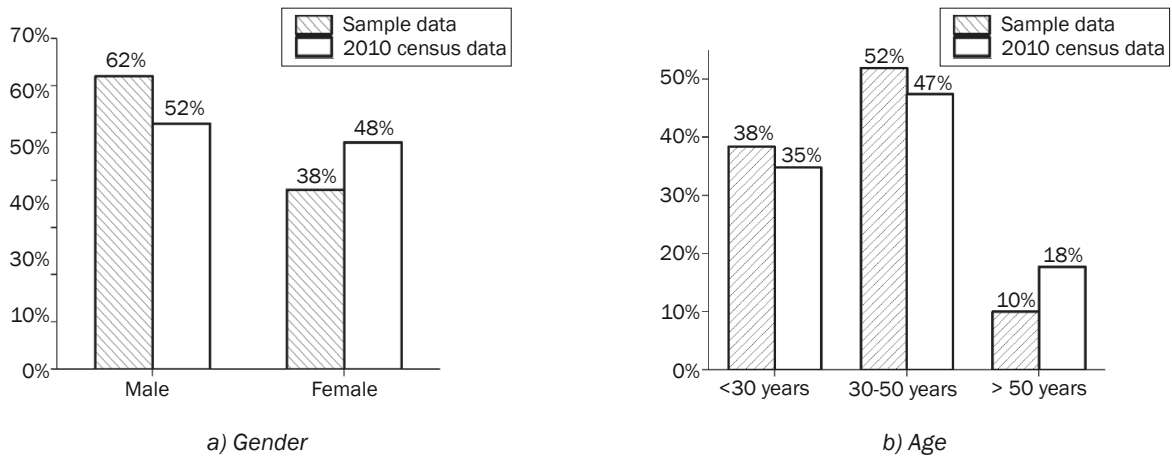


Figure 2 – Socioeconomic variable distribution

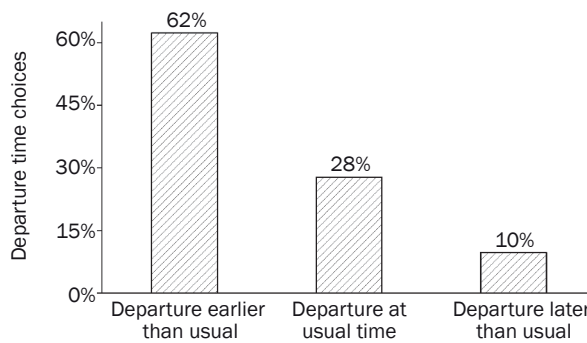


Figure 3 – Distribution of departure time choices

2.2 Data collection and variables definition

The proposed SP survey was conducted in several Beijing subway stations where passenger flow control measures were taken between 7:00 am and 9:00 am from November 2016 to May 2017. The data from a total of 1,860 passengers was collected on weekdays. To verify our sample, we also considered passenger socioeconomic attributes in the survey. Compared with the 2010 census data, there are similar distributions of gender and age, as shown in Figure 2, which proves that the survey data is representative. A summary of departure time choice is illustrated in Figure 3. It reveals passenger willingness to depart earlier, later, or as usual when confronted with fare discount, travel time saving, and less crowding.

The variables of departure time choice are shown in Table 2, and how to model the relationship among these factors will be discussed in the following sections. Note that the code column includes the alternative values of variables.

3. DEPARTURE TIME CHOICE MODELING VIA BAYESIAN NETWORK

The modeling process of departure time choice via BN is given as shown in Figure 4. Enough sample data should be first collected by the survey, as shown in Section 2. Then, the relationships between these

variables are learned through the K2 algorithm [16]. The BN parameters are estimated by the MLE method to calculate the conditional probability distribution.

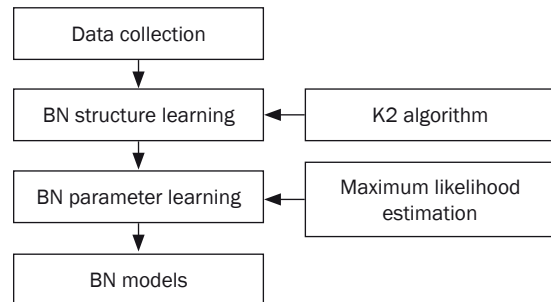






Figure 4 – Flowchart of BN model specification

3.1 BN structure learning

Five nodes are assumed to be independent and created in the BN structure according to the variables mentioned in Section 2.2. The departure time choice node is the final node, and the other four nodes are randomly ranked. Because the node sequence determines the BN structure, we propose two typical sequences for the sake of simplicity: (1) $[X_1, X_2, X_3, X_{14}, X_5]$, and (2) $[X_2, X_3, X_4, X_1, X_5]$. The two associated models are denoted by BN_ONE and BN_TWO, respectively.

Table 2 – Variables description

| Variables | Description | Code |
|---------------------------------|--|--|
| Departure time change (X_1) | Departure 25 minutes earlier | 3 |
| | Departure 20 minutes earlier | 2 |
| | Departure 15 minutes earlier | 1 |
| | Usual departure time | 0 |
| | Departure 15 minutes later | -1 |
| | Departure 20 minutes later | -2 |
| | Departure 25 minutes later | -3 |
| Subway fare (X_2) | Full fare | 1 |
| | 10% discount | 2 |
| | 20% discount | 3 |
| | 30% discount | 4 |
| Travel time saving (X_3) | 0% | 1 |
| | 10% | 2 |
| | 15% | 3 |
| | 20% | 4 |
| Crowding (X_4) | No more than 5 passengers in a queue in front of a train door. |  1 |
| | 6~10 passengers in a queue in front of a train door. |  2 |
| | 11~20 passengers in a queue in front of a train door. |  3 |
| | More than 20 passengers in a queue in front of a train door. |  4 |
| Departure time choice (X_5) | Usual departure time | 1 |
| | Departure earlier than usual | 2 |
| | Departure later than usual | 3 |

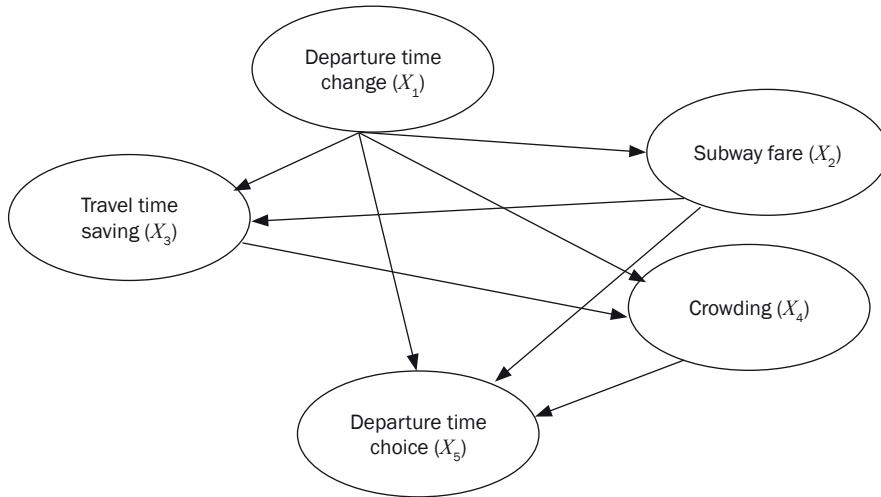


Figure 5 – Structure learning result of BN_ONE

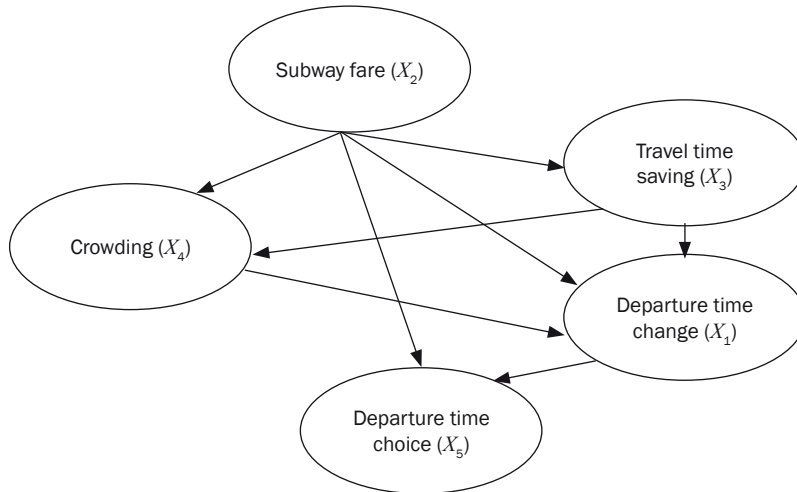


Figure 6 – Structure learning result of BN_TWO

The results of the BN structure are obtained by the K2 algorithm using the BNT toolbox [17], shown in Figure 5 and Figure 6, respectively. Note that the result of BN_ONE indicates that X_5 is dependent on X_1 and X_2 in both BN models, which indicates that departure time change and subway fare are two important factors to decide the departure time choice. In addition, X_5 is directly related to X_4 , unlike in BN_TWO. That is, the structure of BN_ONE considers crowding as a more important factor of departure time choice. Also, X_5 is independent of X_3 in both BN models, which suggests that travel time saving has less impact on departure time choice than the other three factors.

3.2 BN parameter learning

MLE is employed to calculate the conditional probability distribution according to the BN structure. Because all variables are discrete, the conditional probability distribution can be expressed as conditional probability tables (CPT).

One of the CPTs $P(X_5=1 | X_4=1, X_2, X_1)$ for node X_5 in BN_ONE is given in Table 3. Taking $P(X_5=1 | X_4=1, X_2=2, X_1=3)$ for example, the probability of choosing to depart at usual time is 0.6774 under the conditions of first level of crowding, 25 minutes earlier than usual and 10% discount on fare.

Based on the CPTs, the joint probability distribution of decision node in BN_ONE model is computed as follows:

$$\begin{aligned}
 &P(X_1 = x_1, X_2 = x_2, X_4 = x_4, X_5 = x_5) \\
 &= P(X_1 = x_1) \cdot P(X_2 = x_2 | X_1 = x_1) \cdot \\
 &\quad \cdot P(X_4 = x_4 | X_1 = x_1, X_2 = x_2) \cdot \\
 &\quad \cdot P(X_5 = x_5 | X_1 = x_1, X_2 = x_2, X_4 = x_4)
 \end{aligned} \tag{1}$$

Then, the joint probability distribution between X_5 and X_1, X_2, X_4 is given using junction tree algorithm [19] respectively, as follows:

$$\begin{aligned}
 &P(X_5 = x_5, X_1 = x_1) = \\
 &= \sum_{x_2 \in X_2} \sum_{x_4 \in X_4} P(X_5 = x_5, X_1 = x_1, X_2 = x_2, X_4 = x_4)
 \end{aligned} \tag{2}$$

Table 3 – CPT for node X_5 in BN_ONE

| | | | | | |
|---------|----------------------|--------|--------|--------|--------|
| $X_4=1$ | $X_2 \backslash X_1$ | 1 | 2 | 3 | 4 |
| | 3 | 0.0000 | 0.6774 | 0.9871 | 0.0000 |
| | 2 | 0.0000 | 0.0000 | 0.0000 | 0.0000 |
| | 1 | 0.0000 | 0.0645 | 0.0000 | 0.0000 |
| | 0 | 0.0000 | 0.0000 | 0.0000 | 0.0000 |
| | -1 | 0.6774 | 0.0000 | 0.0097 | 0.0000 |
| | -2 | 0.0000 | 0.0516 | 0.8000 | 0.0000 |
| -3 | 0.0000 | 0.0000 | 0.0000 | 0.0000 | |

$$P(X_5 = x_5, X_2 = x_2) = \sum_{x_1 \in X_1} \sum_{x_4 \in X_4} P(X_5 = x_5, X_1 = x_1, X_2 = x_2, X_4 = x_4) \tag{3}$$

$$P(X_5 = x_5, X_4 = x_4) = \sum_{x_1 \in X_1} \sum_{x_2 \in X_2} P(X_5 = x_5, X_1 = x_1, X_2 = x_2, X_4 = x_4) \tag{4}$$

4. RESULT ANALYSIS

4.1 Comparison with other models

To verify the prediction accuracy of the BN models, MNL is developed based on additional observations, and the parameter estimation of travel time saving, crowding, scheduled delay early (SDE), scheduled delay late (SDL), and subway fare in MNL is given in Table 4.

Receiver operating characteristic (ROC) curve is introduced to measure the accuracy, whose key indexes (i.e., false positive rate (FPR) and true positive rate (TPR)) are defined as follows:

$$TPR = \frac{\text{True positive}}{\text{True positive} + \text{False negative}} \tag{5}$$

$$FPR = \frac{\text{False positive}}{\text{False positive} + \text{True negative}} \tag{6}$$

The ROC curve is illustrated in Figure 7, where the area under the curve (AUC) ranges from 0.5 to 1. Specifically, the values of AUC are 0.7480, 0.8203, and 0.8012 for MNL, BN_ONE, and BN_TWO, respectively. The criteria of AUC are displayed in Table 5, and it shows that the three models are accurate in prediction. Usually, a higher AUC value indicates a more accurate prediction. Therefore, BN_ONE and BN_TWO have more reliable prediction accuracy than MNL. Further,

Table 4 – MNL parameter estimation

| Parameters | MNL | T-statistics |
|-----------------|---------|--------------|
| β_{TS} | -0.0461 | -0.74 |
| β_{CROWD} | -1.1621 | -13.26 |
| β_{SDE} | -1.4228 | -20.97 |
| β_{SDL} | -3.2025 | -21.09 |
| β_{FARE} | -1.0396 | -19.30 |

Note: $SDE = \max(\text{usual departure time} - \text{early departure time}, 0)$;
 $SDL = \max(\text{late departure time} - \text{usual departure time}, 0)$.

Table 5 – Criteria of AUC

| AUC | Criteria |
|---------|---------------|
| 0.9~1 | Excellent |
| 0.7~0.9 | Accurate |
| 0.5~0.7 | Less accurate |
| <0.5 | Inaccurate |

BN_ONE performs better than BN_TWO, which shows that the structure of the BN model greatly determines its performance.

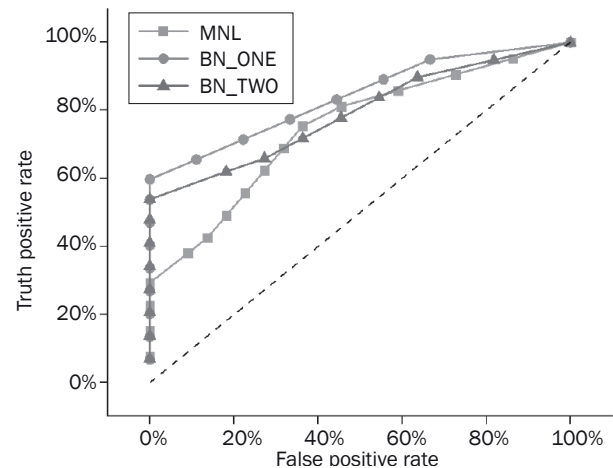


Figure 7 – ROC curve of MNL BN_ONE and BN_TWO

4.2 Analysis of influence on departure time choice

Different from the linear utility function of the discrete choice model, the relationships between two variables can be explained by joint probability in the BN model. Because of its better performance, BN_ONE is employed to analyze the factors of departure time choice.

The joint probability distribution for departure time choice and departure time change is illustrated in Figure 8, and three points are concluded as follows. First, higher probability of choosing usual or earlier departure time can be seen with scenarios of departing 15~25 minutes later (blue, red, and green bars), which indicates strong willingness of subway passengers

to avoid schedule delay [2]. Secondly, if passengers are faced with a scenario of departing 25 minutes early, the probability of choosing early departure time will decrease to 0.4% (grey bar), which reveals that passengers are unwilling to afford extra time cost to depart early. Thirdly, if departure time change is 20 minutes or less, an increase of probability of about 5% for choosing to depart early is observed, which indicates that more passengers will change their departure time by less than 20 minutes. Therefore, motivating passengers to depart early no more than 20 minutes before peak hours is more feasible against departing late.

Regarding subway fare as shown in Figure 9, two points are summarized as follows. First, if passengers are faced with a 30% discount, an increase of probability of about 13.7% (blue bar) for choosing earlier departure time is observed. And if passengers are faced with a 20% discount, an increase of probability of about 9.5% (red bar) is observed, which indicates that higher fare discount will attract passengers to

depart earlier. Second, if passengers enjoy a 10% or lower discount on subway fare, the probability of choosing to depart earlier decreases (green and purple bars), which reveals that less than 10% discount is not attractive. Hence, a fare discount deeper than 20% before peak hours will motivate passengers to depart earlier, and trip cost has been proven as an important factor for departure time choice behaviors [18].

As for crowding shown in Figure 10, two points are concluded as follows. First, if passengers are faced with a scenario with the first level of crowding, an increase of the probability of choosing to depart earlier by about 5.0% (blue bar) is observed. If passengers are faced with the second level of crowding, the probability for choosing to depart at usual time increases by about 6.6% (red bar). These are due to the fact that less crowding means getting a chance to have a seat [19]. Second, if the crowding becomes serious, about 10.0% increase of probability (green and purple bar) for choosing earlier departure time is observed, which reveals that serious crowding has an influence on

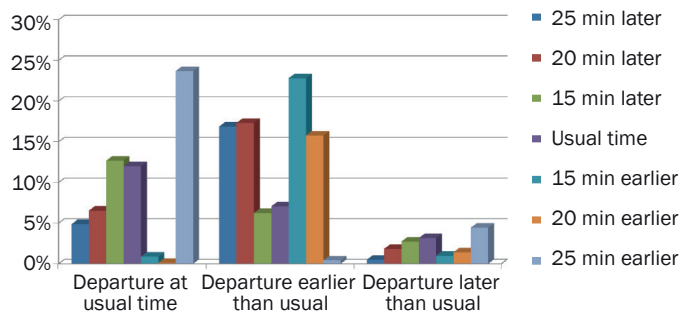


Figure 8 – Joint probability distribution of departure time choice and departure time change

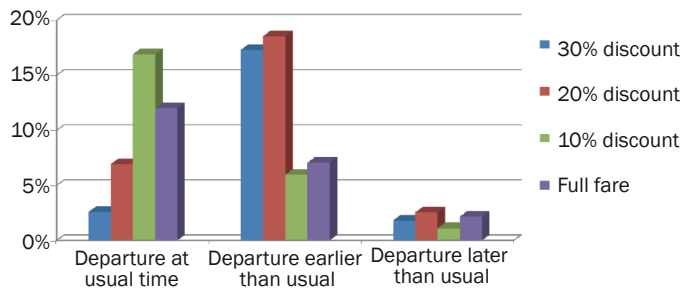


Figure 9 – Joint probability distribution of departure time choice and subway fare

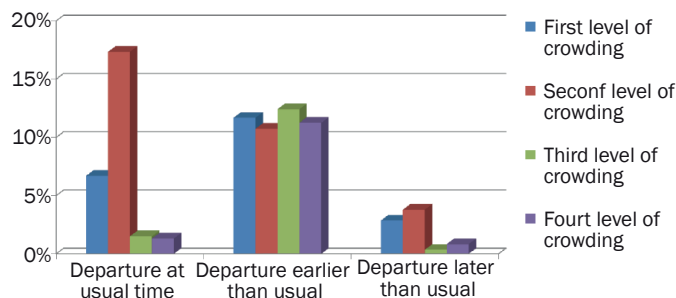


Figure 10 - Joint probability distribution of departure time choice and crowding

arrival time, and passengers have to depart earlier than usual. This proves that serious crowding during peak hours is a motivating factor for passengers to depart earlier.

5. CONCLUSIONS

In this study, BN models are proposed to model departure time choice of subway passengers. With the help of SP data collected in subway stations, the K2 algorithm and MLE are employed to estimate the BN structure and its parameters, respectively. An MNL model is developed to verify the accuracy of the proposed model. Finally, the relationships between departure time choice and other factors are discussed.

The main findings are summarized as follows:

- 1) BN models present a distinct advantage over the MNL model in terms of prediction accuracy. Therefore, BN is a powerful method for inference of subway passengers' departure time choice.
- 2) Further, the relationships of probability change between departure time choice and attributes can be analyzed through BN, which is free from linear constraints in the utility function. Under the conditions of more than 20% discount on subway fare, less than 20 minutes of departure time change, and less serious crowding, a higher probability of choosing earlier departure time is observed.

Unlike departure time change and crowding, providing more than 20% discount on subway fare is an easier and more practical way to guide passengers to depart earlier than usual and alleviate the pressure of crowding in stations. These findings will contribute to the creation of passenger flow control strategies and passenger volume prediction in subway operations.

In the future studies, other factors, such as travel experience, seasons, weather, and incidents, will also be explored.

ACKNOWLEDGMENT

The authors gratefully acknowledge the support provided by the National Natural Science Foundation of China (No.71601018), State Key Lab of Rail Traffic Control and Safety, China (RCS2017ZT003), the Fundamental Research Funds for the Central Universities (2017JBM075), China, and Social Science Foundation of Jiangsu Province (Grant no. 14EYD010).

李宪, 硕士研究生^{1,2}

邮箱: 15120846@bjtu.edu.cn

李海鹰, 教授, 博士¹

邮箱: hyl@bjtu.edu.cn

许心越, 副教授, 博士¹

(通讯作者)

邮箱: xxy@bjtu.edu.cn

¹ 北京交通大学 轨道交通控制与安全国家重点实验室
北京, 中国 100044

² 北京交通大学 交通运输学院
北京, 中国 100044

基于贝叶斯网络的出行时间选择建模-以北京地铁为例

摘要

出行时间选择对于地铁乘客避开高峰时段拥堵至关重要, 本文提出了一种基于贝叶斯网络的出行时间选择模型构建方法。首先, 模型考虑了旅行时间节省、拥挤度、地铁票价和出行时间调整量四个因素; 进而结合收集的SP调查数据, 采用K2算法进行贝叶斯网络的结构学习, 并利用极大似然估计进行参数学习; 最后, 以北京地铁为案例, 验证了本文提出的贝叶斯网络模型比典型的离散选择模型具有更高的预测精度; 并对票价的敏感性进行了分析, 进一步为地铁管理者提供了客流组织建议: 八折及以上的票价优惠将有利于促使乘客提前15~20分钟提前出行, 以减缓早高峰的车站拥堵压力。

关键词

出行时间选择; 贝叶斯网络; 拥堵; 地铁乘客

REFERENCES

- [1] Xu X, Liu J, Li H, et al. Capacity-oriented passenger flow control under uncertain demand: Algorithm development and real-world case study. *Transportation Research Part E: Logistics and Transportation Review*. 2016;87: 130-148.
- [2] Jong GD, Daly A, Pieters M, et al. A model for time of day and mode choice using error components logit. *Transportation Research Part E: Logistics and Transportation Review*. 2003;39(3): 245-268.
- [3] Thorhauge M, Haustein S, Cherchi E. Accounting for the Theory of Planned Behaviour in departure time choice. *Transportation Research Part F: Traffic Psychology and Behaviour*. 2016;38: 94-105.
- [4] Habib KN, Day N, Miller EJ. An investigation of commuting trip timing and mode choice in the greater Toronto area: Application of a joint discrete-continuous model. *Transportation Research Part A: Policy and Practice*. 2009;43(7): 639-653.
- [5] Hess S, Daly A, Rohr C, et al. On the development of time period and mode choice models for use in large scale modelling forecasting systems. *Transportation Research Part A: Policy and Practice*. 2007;41(9): 802-826.
- [6] Bajwa SU, Bekhor S, Kuwahara M, et al. Discrete choice modeling of combined mode and departure time. *Transportmetrica*; 2008;4(2): 155-177.
- [7] Sasic A, Habib KN. Modelling departure time choices by a Heteroskedastic Generalized Logit (Het-GenL) model: An investigation on home-based commuting trips in the Greater Toronto and Hamilton Area (GTHA). *Transportation Research Part A: Policy and Practice*, 2013;50(2): 15-32.
- [8] Jou R. Modeling the impact of pre-trip information on commuter departure time and route choice. *Transportation Research Part B: Methodological*. 2001;35(10): 887-902.
- [9] Schwanen T, Ettema D. Coping with unreliable transportation when collecting children: Examining parents' behavior with cumulative prospect theory. *Transportation Research Part A: Policy and Practice*. 2009;43(5): 511-525.

- [10] Chorus CG, Arentze T, Timmermans H, et al. A random regret minimization model of travel choice. *Transportation Research Part B: Methodological*. 2008;42(1): 1-18.
- [11] Zhu Z, Chen X, Xiong C, et al. A mixed Bayesian network for two-dimensional decision modeling of departure time and mode choice. *Transportation*. 2017: 1-24.
- [12] Zhang K, Taylor MAP. Effective arterial road incident detection: A Bayesian network based algorithm. *Transportation Research Part C: Emerging Technologies*. 2006;14(6): 403-417.
- [13] Nozick LK, Xie C, Wang H. Modeling Travel Mode Choice Behavior by Bayesian Network. *Transportation Research Board 85th Annual Meeting, 22-26 Jan 2006, Washington DC, USA*; 2006.
- [14] Gao JX, Zhi-Cai J, An-Ning NI. Modeling and Applications of Traveler Destination Choice Behavior based on Bayesian Network. *Journal of Systems & Management*. 2015;108(2): 289-295.
- [15] Huber J, Zwerina K. The Importance of Utility Balance in Efficient Choice Designs. *Journal of Marketing Research*. 1996;33(3): 307-317.
- [16] Cooper GF, Herskovits EH. A Bayesian Method for the Induction of Probabilistic Networks from Data. *Machine Learning*. 1992;9(4): 309-347.
- [17] Murphy KP. *The Bayes Net Toolbox for Matlab*. Computing Science & Statistics; 2001.
- [18] Saleh W, Farrell S. Implications of congestion charging for departure time choice: Work and non-work schedule flexibility. *Transportation Research Part A: Policy and Practice*. 2005;39(7-9): 773-791.
- [19] Tirachini A, Hensher DA, Rose JM, et al. Crowding in public transport systems: effects on users, operation and implications for the estimation of demand. *Transportation Research Part A: Policy and Practice*. 2013;53: 36-52.