

基于信息熵模型的地铁乘客出行规律稳定性度量^{*}

洪玲^{1,2} 原梦^{1,2} 刘伟³ 彭秀秀⁴ 江志彬^{1,2}

(1. 同济大学道路与交通工程教育部重点实验室, 201804, 上海;

2. 上海市轨道交通结构耐久与系统安全重点实验室, 201804, 上海;

3. 上海申通地铁集团有限公司技术中心, 201103, 上海; 4. 上海市隧道工程轨道交通设计研究院, 200235, 上海)

摘要 [目的] 为更精确地度量城市轨道交通乘客个体出行规律稳定性, 提升出行预测精度, 解决现有信息熵模型在量化稳定性时存在的偏差问题, 为交通规划和管理提供更可靠的数据支持。[方法] 首先, 构建以单日出行站点次序为基础的日出行链, 作为分析乘客出行规律的核心数据结构。其次, 基于 LCS(最长公共子序列) 算法设计日出行链相似性度量方法, 通过计算日出行链之间的 LCS 长度评估其相似性, 并构建相似度矩阵以表征个体出行的日间重复程度。为进一步提高稳定性度量的准确性, 引入修正系数 α 对传统信息熵模型进行改进, 以消除基础模型输出熵值中的偏离。最后, 通过实际乘客出行案例, 构建乘客日出行链, 分别计算修正前后的信息熵值, 对比分析模型改进效果, 并验证其不同乘客群体中的适用性。[结果及结论] 改进后的信息熵模型能有效消除基础模型输出熵值中的偏离, 对乘客出行链规律程度的量化输出更符合实际出行的稳定性特征, 对多数乘客有较好的适应性。该方法为城市轨道交通乘客出行规律的精确度量提供了新的思路, 有助于提升出行预测的精度和可靠性, 为交通运营优化提供科学依据。

关键词 城市轨道交通; 乘客出行规律; 日出行链; 出行规律稳定性; 信息熵

中图分类号 U293.13

DOI:10.16037/j.1007-869x.2025.03.017

Stability Measurement of Metro Passenger Travel Patterns Based on Information Entropy Model

HONG Ling^{1,2}, YUAN Meng^{1,2}, LIU Wei³, PENG Xiuxiu⁴, JIANG Zhibin^{1,2}

(1. Key Laboratory of Road and Traffic Engineering of the Ministry of Education, Tongji University, 201804, Shanghai, China; 2. Shanghai Key Laboratory of Structural Durability and System Safety for Railway Transportation, 201804, Shanghai, China; 3. Technical Center of Shanghai Shentong Metro Group Co., Ltd., 201103, Shanghai, China; 4. Shanghai Tunnel Engineering & Rail Transit Design and Re-

search Institute, 200235, Shanghai, China)

Abstract [Objective] It is aimed to more accurately measure the individual travel pattern stability of urban rail transit passengers, improve travel prediction accuracy, and address bias issues in existing information entropy models when quantifying stability, providing more reliable data support for transportation planning and management. [Method] First, a daily travel chain is constructed based on the sequence of stations visited in a single day, serving as the core data structure for analyzing passenger travel patterns. Next, a similarity measurement method for daily travel chains is designed based on LCS (longest common subsequence) algorithm. The similarity between daily travel chains is evaluated by calculating their LCS length, and a similarity matrix is constructed to present the daily travel repetition degree for individuals. To further enhance the accuracy of stability measurement, a correction coefficient α is introduced to improve the traditional information entropy model, eliminating deviations in the entropy values produced by the base model. Finally, real passenger travel cases are used to construct daily travel chains, and the information entropy values before and after correction are calculated. A comparative analysis is conducted to assess the effectiveness of the model improvements and verify its applicability to different passenger groups. [Result & Conclusion] The improved information entropy model effectively eliminates deviations in the entropy values of the base model, providing a more accurate quantification output of the passenger travel pattern stability, adequately adaptable to most passengers. This method offers a new approach to accurately measuring urban rail transit passenger travel patterns, helping to improve travel prediction accuracy and reliability while providing a scientific basis for optimizing transportation operations.

Key words urban rail transit; passenger travel pattern; day trip chain; travel pattern stability; information entropy

^{*} 国家自然科学基金项目(52372332); 上海市隧道工程轨道交通设计研究院科研计划项目(Y202431)

个体乘客的出行行为预测是城市轨道交通客流精细化运营管理的重要依据,而因为不同乘客出行行为间的异质性,乘客出行过程预测需要考虑乘客出行规律的稳定程度。信息熵方法常被用于度量样本的规律程度,在原理上也同样适用于个体出行规律性度量。通过得到个体乘客出行规律稳定性指标,可以提升个体出行预测精度。

在基于信息熵模型度量乘客出行规律稳定程度的相关研究中,文献[1-4]提出并推广了使用出行序列分布估算得到的熵率来表征乘客出行规律稳定性的方法。在得到出行序列分布的过程中,文献[5]将相同站点集群的一日出行链归为同一类出行模式;文献[6]则通过识别主要活动区域进行出行链集合的简化提取。

信息熵模型适用于对一个离散化概率分布的信息量测度,已有研究一般通过语义识别等方式将原始出行行为处理为概率分布,这类处理方式会忽略乘客出行的微小波动,此时由日间出行相似性带来的预期外信息累加可能使得模型的熵值输出偏离实际的出行稳定性水平。因此需要设计一种能精确体现原始进出站空间信息规律的稳定性度量方法。

1 出行规律稳定性度量方法

信息熵模型的优点在于其具有单调性与累加性。单调性使信息熵指标能够支撑不同乘客的出行规律强度对比,即熵值越小,出行规律性越强;累加性使模型解释性更强,即新出行模式出现时会给系统熵值增加非负数,系统累积的信息熵值变大。

1.1 数据来源与预处理

AFC(自动售检票)系统数据能记录乘客在城市轨道交通网络中的进出站行为,采用 AFC 收集的信息包含卡号、进出站点、进出站时间等乘客出行关键信息,如表 1 所示。表中 A—F 表示进出车站站点。

表 1 AFC 数据字段
Tab.1 AFC data fields

| 卡号 | 进站车站 | 进站时间 | 出行日期 | 出站车站 | 出站时间 |
|---------|------|-------|------------|------|-------|
| 8 * * 3 | A | 13:06 | 2021-07-13 | B | 13:44 |
| 4 * * E | C | 07:49 | 2021-07-13 | D | 08:40 |
| E * * 8 | E | 07:39 | 2021-07-13 | F | 08:17 |

基于 AFC 数据,根据单个乘客在运营日内的进

出站行为集合,按照乘客行为次序,构建出如表 2 所示的完整日出行链:站点代号 A、B、C、D 表示进站点,站点代号 -A、-B、-C、-D 表示出站点,数字表示对应进出站行为的发生次序。

表 2 日出行链构建示例

Tab.2 Example of daily travel chain construction

| 出行日期 | 不同次序的进出站行为 | | | | | |
|----------|------------|----|---|----|---|----|
| | 1 | 2 | 3 | 4 | 5 | 6 |
| 20210801 | A | -B | B | -A | \ | \ |
| 20210802 | A | -C | C | -A | \ | \ |
| 20210803 | A | -B | B | -D | C | -A |

1.2 基础信息熵模型

在使用基础信息熵模型对乘客进行规律性度量时,以表 2 的处理方式得到其在指定日期段 T (天数)内的出行链,形成出行链集合 R 。

对于集合 R 可以用离散随机变量 X 代表该乘客的出行链分布, $p(x)$ 代表 X 中某一出行链 x 的发生概率,因此离散变量系统 X 的信息熵能反映乘客出行链集合的规律程度。 X 的规律性越强,其信息熵 $H(X)$ 就越低,反之规律性越弱则信息熵越高。信息熵 $H(X)$ 模型计算公式为:

$$H(X) = - \sum_{x \in R} p(x) \lg p(x) \quad (1)$$

单个事件 x 的信息熵 $h(x)$ 由其概率 $p(x)$ 决定(见式(2)),多个事件 x 的信息熵取和为系统 X 的总信息熵(见式(3))。信息熵模型计算公式(1)可分别表示为下式:

$$h(x) = - p(x) \lg p(x) \quad (2)$$

$$H(X) = \sum_{x \in R} h(x) \quad (3)$$

在使用信息熵模型计算出行规律稳定性时,乘客所有出行模式的信息熵值的加和结果即为评估其整体出行规律稳定性的总信息熵值。信息熵模型有其应用条件:对于 X 的事件集 R ,其中所包含的事件 x 需互相独立,互相之间不存在信息重合。该条件在乘客出行链场景下的描述为:对单个乘客而言,其不同出行链模式之间互相独立,不存在部分相似程度。

2 信息熵模型改进

实际中存在多数乘客不同日间的出行行为,包含部分重合的情况,不满足式(3)应用的独立性要求,因此需对基础信息熵模型进行改进。

2.1 修正系数 α 引入

根据事件 x 与其他事件的相似程度引入修正系数 α 对其信息熵 $h(x)$ 进行修正,改进后的信息熵 $h'(x)$ 的计算公式为:

$$h'(x) = -\alpha p(x) \lg p(x) \quad (4)$$

通过引入 α ,实现对事件 x 的信息熵值进行削减修正,与其他事件的相似程度越高, α 越小, $h(x)$ 的损失程度越高。在出行规律稳定性评估场景中,事件 x 等价于乘客在一天内的出行序列,采用 LCS (最长公共子序列) 算法^[7-8] 来实现 α 的评估。LCS 算法能识别并提取两个出行序列之间共有的、分散在不同位置的相同子序列。例如乘客在原有出行链中途增加额外接驳点时,LCS 算法能够保留出行序列中未变的部分。

2.2 修正系数 α 计算方式

对于含有 k 个事件的出行链集合 R ,其中某出行链 x 的修正系数 α_i 具体的计算步骤如下:

步骤1 计算 R 中两条日出行链 x_i 和 x_j 之间的相似程度 $s(x_i, x_j)$ 。 $l(x_i, x_j)$ 表示两日出行间的相似部分长度,与原有两序列长度 $l(x_i)$ 和 $l(x_j)$ 平均值之比可以表征相似度 $s(x_i, x_j)$,计算公式为:

$$s(x_i, x_j) = \frac{2l(x_i, x_j)}{l(x_i) + l(x_j)} \quad (5)$$

步骤2 构建修正系数矩阵 A 。 A 中元素 $A(i, j)$ 表示事件 x_j 对 x_i 造成影响后 x_i 的信息留存比例,影响 $A(i, j)$ 的因素包括 x_i 的概率 $p(x_i)$ 、 x_j 的概率 $p(x_j)$ 、 x_i 与 x_j 的相似度 $s(x_i, x_j)$ 。考虑事件 x_j 的影响后, x_i 的实际信息量 $H'(x)$ 计算公式为:

$$H'(X) = A(i, j)H(X) \quad (6)$$

依照信息熵模型中对事件信息量的定义,可得到 $A(i, j)$ 的计算式:

$$A(i, j) = 1 - s(x_i, x_j) \left(1 - \frac{\lg(p(x_i) + p(x_j))}{\lg p(x_i)} \right) \quad (7)$$

步骤3 计算修正系数 α_i 。对于 x_i ,其信息量变化受其他 $k-1$ 个出行链的共同影响,即矩阵 A 第 i 行所有元素的影响。单个元素对 x_i 的影响如式(7), $k-1$ 个元素的共同影响下, α_i 的计算公式为:

$$\alpha_i = \prod_{j=1}^{k-1} A(i, j) \quad (8)$$

综上所述,引入修正系数 α 后,系统 X 的修正信息熵值计算公式为:

$$H'(X) = \sum_{x_i \in R} \alpha_i h(x_i) \quad (9)$$

3 合理性验证

对修正后的信息熵模型,需通过实际乘客出行案例验证模型改进前后的效果。

3.1 乘客日出行链构建

以上海轨道交通网络中的两名乘客 M、N 为例,其单日出行链集合分别如表3和表4所示。乘客 M 的4种出行链间存在高度的相似重合度,乘客 N 的5种出行链间的相似重合度较低。同时乘客 M、N 的日出行链有着较为接近的占比分布。

表3 乘客 M 出行链集合及占比

Tab.3 Aggregation and share of passenger M travel chains

| 日出行链 | 占比/% |
|--------------|------|
| A, -B, B, -A | 57.1 |
| C, -B, B, -C | 23.8 |
| C, -B, B, -A | 14.3 |
| A, -B, B, -C | 4.8 |

表4 乘客 N 出行链集合及占比

Tab.4 Aggregation and share of passenger N travel chains

| 日出行链 | 占比/% |
|--------------|------|
| A, -B | 50.0 |
| B, -C | 22.7 |
| B, -D | 18.2 |
| E, -F, F, -C | 4.5 |
| G, -H | 4.5 |

M、N 两名乘客在 2021 年 8 月份工作日的出行记录分别如图1和图2所示。图中黑色代表乘客的地铁出行,其他颜色代表乘客停留的实际区域,横向序列从左到右的颜色变化即代表乘客在一天中停留区域的变化,乘客所有颜色序列之间的相似性就反映了乘客出行模式的稳定性。

M 乘客不同出行链间有更多重合,停留区域相对固定,日间出行关联程度较高;N 乘客停留区域变化较大,日间出行关联程度较低。由图表信息可知 M 的空间活动稳定程度显著高于 N。

3.2 模型改进效果

针对乘客 M、N,依照日出行链类型分别计算模型修正前后的信息熵值,结果如表5所示。经对比可得,使用基础信息熵模型进行稳定性度量时,乘客 M、N 的原输出熵值分别为 1.56、1.83,较为接近;使用改进后模型得到乘客 M、N 的修正熵值分

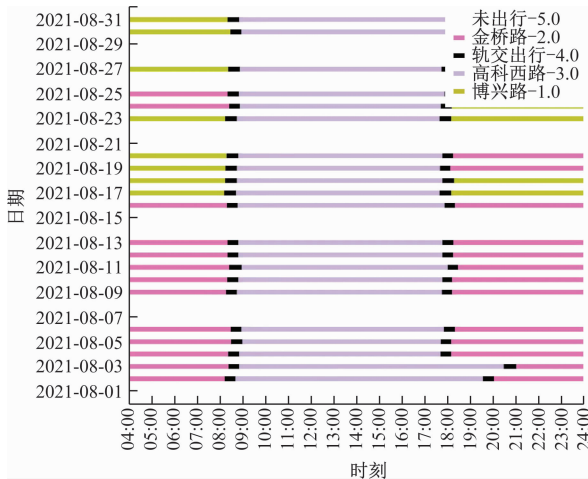


图1 乘客M出行记录

Fig. 1 Passenger M travel records

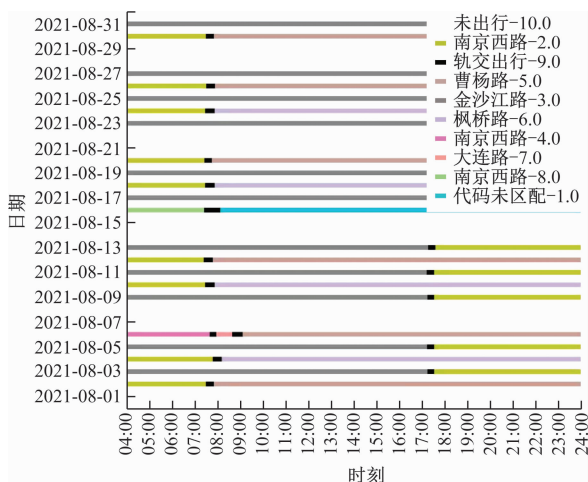


图2 乘客N出行记录

Fig. 2 Passenger N travel records

别为0.51、1.58,改进后模型能区分乘客M和N之间的稳定性强度差异。

将乘客M和N的信息熵值置于整体乘客群体的分布中,分析乘客信息熵值的相对大小。由1000名乘客作为整体样本数据,绘制出如图3所示的所有乘客信息熵值在修正前后的变化情况,其中红色表示基础信息熵值的分布情况,蓝色表示模型改进后的修正信息熵值。由图3可见,修正模型对乘客整体的熵值分布有明显的改变效果。

由图3可知,改进前乘客M、N在乘客群体中的分布位置较为接近,改进后分布位置差异较大。改进后的模型消除了日出行链间的重合相似部分,能说明M、N两乘客的规律性差异水平。

对单个乘客而言,修正模型产生的熵值削减部分可以体现修正模型对该乘客的有效性程度,计算

表5 模型改进前后乘客M、N出行稳定性度量结果

Tab. 5 Metric results of travel stability for passengers M and N before and after model improvements

| 乘客ID | 日出行链 | 占比/% | 原输出熵值 | 修正熵值 |
|------|--------------|--------|-------|------|
| M | A, -B, B, -A | 57.1 | 0.46 | 0.20 |
| | C, -B, B, -C | 23.8 | 0.49 | 0.19 |
| | C, -B, B, -A | 14.3 | 0.40 | 0.08 |
| | A, -B, B, -C | 4.8 | 0.20 | 0.03 |
| | 合计 | 100.00 | 1.56 | 0.51 |
| N | A, -B | 50.0 | 0.50 | 0.50 |
| | B, -C | 22.7 | 0.48 | 0.37 |
| | B, -D | 18.2 | 0.44 | 0.34 |
| | E, -F, F, -C | 4.6 | 0.20 | 0.16 |
| | G, -H | 4.5 | 0.20 | 0.20 |
| | 合计 | 100.00 | 1.83 | 1.58 |

得到不同修正效用程度的乘客占比如表6所示。在使用基础模型进行出行稳定性度量时,仅有6.1%的乘客修正效用小于30%,而9.8%的乘客修正效果超过70%。

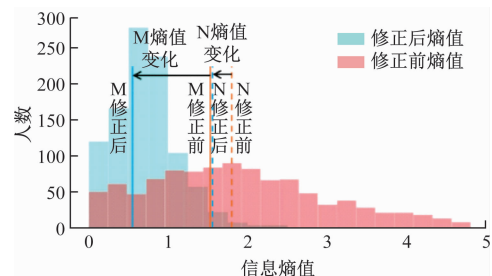


图3 模型改进前后乘客M、N的相对信息熵值

Fig. 3 Relative information entropy values of passengers M and N before and after model improvement

表6 各修正效用程度的乘客占比

Tab. 6 Proportion of passengers by degree of correction effectiveness

| 修正效用程度 | 乘客占比/% |
|--------------|--------|
| 低(<30%) | 6.1 |
| 中(30%~<50%) | 18.5 |
| 较高(50%~<70%) | 65.6 |
| 高(≥70%) | 9.8 |

4 结语

为了提升对个体出行稳定性的度量精度,本文构建了以单日出行站点次序为依据的日出行链,提出了一种信息熵模型的改进方法。通过引入对各

出行链输出熵值的修正系数,并以链间相似度指标确定系数,使改进后的模型有更精确的规律性水平量化结果,适用的乘客范围更广。但在熵值合理性验证方面,目前主要从乘客出行链与乘客在群体中的相对规律性水平进行考量,仍需要进一步探索能证明输出熵值合理性的方法。

参考文献

- [1] GOULET-LANGLOIS G, KOUTSOPOULOS H N, ZHAO J. Inferring patterns in the multi-week activity sequences of public transport users[J]. *Transportation Research Part C: Emerging Technologies*, 2016, 64: 1.
- [2] GOULET-LANGLOIS G, KOUTSOPOULOS H N, ZHAO Z, et al. Measuring regularity of individual travel patterns[J]. *IEEE Transactions on Intelligent Transportation Systems*, 2018, 19(5): 1583.
- [3] LEI D, CHEN X, CHENG L, et al. Inferring temporal motifs for travel pattern analysis using large scale smart card data[J]. *Transportation Research Part C: Emerging Technologies*, 2020, 120: 102810.
- [4] LEI D, CHEN X, CHENG L, et al. Minimum entropy rate-improved trip-chain method for origin-destination estimation using smart card data[J]. *Transportation Research Part C: Emerging Technologies*, 2021, 130: 103307.
- [5] 姚志刚, 杨杰, 王元庆. 基于个体出行模式的公交乘客活动

规律性度量[J]. *北京交通大学学报*, 2022, 46(4): 68.

YAO Zhigang, YANG Jie, WANG Yuanqing. Measurement of public transport passenger behavior regularity based on individual travel pattern[J]. *Journal of Beijing Jiaotong University*, 2022, 46(4): 68.

- [6] 张娜, 陈峰, 王剑坡, 等. 基于时空序列相似性的城轨乘客出行模式识别[J]. *吉林大学学报(工学版)*, 2024, (9): 2588.
- ZHANG Na, CHEN Feng, WANG Jianbo, et al. Recognition of travel patterns for urban rail transit passengers based on spatiotemporal sequence similarity[J]. *Journal of Jilin University (Engineering and Technology Edition)*, 2024, (9): 2588.
- [7] SHOU Z, DI X. Similarity analysis of frequent sequential activity pattern mining[J]. *Transportation Research Part C: Emerging Technologies*, 2018, 96: 122.
- [8] FAROQI H, MESBAH M, KIM J. Spatial-temporal similarity correlation between public transit passengers using smart card data[J]. *Journal of Advanced Transportation*, 2017, 2017: 1.

· 收稿日期:2024-08-25 修回日期:2024-09-30 出版日期:2025-03-10
Received:2024-08-25 Revised:2024-09-30 Published:2025-03-10
· 第一作者:洪玲, 正高级工程师, ss980308@tongji.edu.cn
通信作者:江志彬, 副教授, jzb@tongji.edu.cn
· ©《城市轨道交通研究》杂志社, 开放获取 CC BY-NC-ND 协议
© Urban Mass Transit Magazine Press. This is an open access article under the CC BY-NC-ND license

(上接第92页)

- [7] 姜丽娟. S01-10 型蓄电池牵引车[J]. *工程机械*, 2005, 36(10): 12.
- JIANG Lijuan. Model S01-10 battery powered tractor[J]. *Construction Machinery and Equipment*, 2005, 36(10): 12.
- [8] 赵辉, 李铁才, 孙立志, 等. 电池供电的永磁电动机系统的再生制动[J]. *电机与控制学报*, 1999, 3(4): 207.
- ZHAO Hui, LI Tiecai, SUN Lizhi, et al. Regenerative braking of permanent magnet motor system fed by battery[J]. *Electric Machines and Control*, 1999, 3(4): 207.
- [9] 张传林, 胡文静. 稀土永磁材料的发展及在电机中的应用[J]. *微电机*, 2003, 36(1): 38.
- ZHANG Chuanlin, HU Wenjing. The development of the rare-earth material and its applications to the electric machine[J]. *Micromotors*, 2003, 36(1): 38.
- [10] 裴建红. 基于碳化硅及永磁电机技术的新型城轨车辆牵引系统研究[J]. *江苏科技信息*, 2020, 37(8): 41.
- PEI Jianhong. Research on new traction system of urban rail vehicle based on silicon carbide and permanent magnet motor technology[J]. *Jiangsu Science & Technology Information*, 2020, 37

(8): 41.

- [11] 唐朝辉, 唐立国, 李宝泉, 等. 地铁车辆永磁牵引系统与异步牵引系统的对比分析[J]. *电工技术*, 2018(24): 142.
- TANG Zhaohui, TANG Liguang, LI Baoquan, et al. Comparative analysis of permanent magnet traction system and asynchronous traction system for metro vehicles[J]. *Electric Engineering*, 2018(24): 142.
- [12] 朱伟鹏. 深圳地铁永磁同步牵引系统研究分析[J]. *现代城市轨道交通*, 2021(4): 38.
- ZHU Weipeng. Research and analysis of permanent magnet synchronous traction system for Shenzhen Metro[J]. *Modern Urban Transit*, 2021(4): 38.

· 收稿日期:2023-03-07 修回日期:2023-05-20 出版日期:2025-03-10
Received:2023-03-07 Revised:2023-05-20 Published:2025-03-10
· 第一作者:周成尧, 高级工程师, 13811491797@163.com
通信作者:赵雷廷, 副研究员, zhaoleiting@zemt.cn
· ©《城市轨道交通研究》杂志社, 开放获取 CC BY-NC-ND 协议
© Urban Mass Transit Magazine Press. This is an open access article under the CC BY-NC-ND license