

基于随机森林——隐马尔可夫融合算法的 乘客轨迹还原与意图识别

贾小艺

上海理工大学管理学院, 上海

收稿日期: 2026年3月8日; 录用日期: 2026年4月1日; 发布日期: 2026年4月8日

摘要

随着城市轨道交通网络的快速扩张, 地铁站作为客流集散的关键节点, 其设施布局的合理性直接关系到运营效率与乘客满意度。然而, 现有的车站设施布局往往依赖于设计经验及行业规范, 缺乏对乘客出行意愿的挖掘, 易出现动线冲突与高峰时段局部拥堵。针对这一问题, 本文提出了一种融合隐马尔可夫模型与随机森林的乘客行为推演模型, 构建了考虑乘客出行意愿的动线预测模型, 丰富了地铁车站设施布局优化的理论依据, 为运营部门进行车站局部改造及客流组织优化提供了可操作的决策支持。

关键词

地铁站, 乘客轨迹还原, 隐马尔可夫模型, 随机森林

Passenger Trajectory Reconstruction and Intention Recognition Based on Random Forest-Hidden Markov Fusion Algorithm

Xiaoyi Jia

Business School, University of Shanghai for Science & Technology, Shanghai

Received: March 8, 2026; accepted: April 1, 2026; published: April 8, 2026

Abstract

With the rapid expansion of urban rail transit networks, subway stations serve as critical hubs for passenger flow distribution. The rationality of their facility layout directly impacts operational efficiency and passenger satisfaction. However, existing station layouts often rely on design experience and industry standards, lacking consideration of passenger travel preferences. This frequently

leads to conflicts in movement paths and localized congestion during peak hours. To address this issue, this paper proposes a passenger behavior simulation model integrating Hidden Markov Models (HMMs) and Random Forests. It constructs a movement path prediction model that incorporates passenger emergence preferences, thereby enriching the theoretical basis for optimizing subway station facility layouts. This approach provides actionable decision support for operational departments undertaking localized station renovations and optimizing passenger flow management.

Keywords

Subway Station, Passenger Trajectory Reconstruction, Hidden Markov Model, Random Forest

Copyright © 2026 by author(s) and Hans Publishers Inc.

This work is licensed under the Creative Commons Attribution International License (CC BY 4.0).

<http://creativecommons.org/licenses/by/4.0/>



Open Access

1. 引言

自二十世纪八十年代新中国第一条地铁交付运营以来，我国城市轨道交通经历了从无到有、从单线运营到网络化布局的跨越式发展。根据交通运输部最新数据统计，截止 2024 年底，我国内地共有 58 个城市开通运营城市轨道交通线路 361 条，运营里程 12160.77 公里，稳居世界领先地位。与此同时，多地政府通过出台政策补贴[1]，进一步推动了地铁出行的普及，使其成为了越来越多市民出行的首选交通方式。然而，客流的快速增长也给运营管理带来了巨大压力[2]，如何在高负荷下保持高质量运营成为了亟待解决的问题。未来，我国轨道交通运营也将从“增量发展为先”的大规模发展向以“存量维护为主”的精细化生产时代转变。在这一背景下，如何优化运营效能、提升乘客满意度也将成为行业关注的焦点。

本文聚焦于“设施布局”这一影响管理效能的基础物理载体，提出一个基于数据驱动与系统规划相结合的针对性研究框架。通过对真实的乘客行为数据分析，预测并识别不同出行意志下乘客对实时及动线选择的关系，以此为指导，对地铁站内高频设施进行布局优化，通过提升车站运营的效率、减少设施闲置带来的资源浪费，进而提升服务水平及乘客满意度。首先，整合地铁站内 AFC 系统数据、视频监控录像及等多源异构信息，构建乘客在进入地铁站系统后、进入列车之前的动线时空序列。其次，针对乘客动线的随机性与隐含状态的关联性，引入隐马尔可夫模型(HMM)进行动线建模。最后，将 HMM 提取的特征与乘客属性、时间上下文等特征相结合，构建随机森林分类与回归模型，遵循“数据采集与处理→行为状态识别→高频动线预测”的逻辑主线，为后续识别高频、关键动线奠定坚实的数据分析基础。

2. 文献综述

当前，国内关于轨道交通管理效能的优化与提升主要聚焦于服务质量、风险管理、资产管理、网络化布局、能耗需求等几个方面：

服务质量。国内对于服务质量的界定还未形成明确清晰的界定[3]。一些学者聚焦到被服务的微观个体，侧重于用公众满意度来反映公共服务质量[4]；另一派学者以更宏观的视角评价服务水平，更倾向于关注轨道交通服务的普遍性[3] [5]、非营利性、抗风险能力[6]等关键特征。

风险管理。影响地铁系统安全的风险因素可以分为四大类[7]：人员因素、机械因素、环境因素、管理因素。人员因素包括：人员的身心健康状况、专业培训情况、文化程度、安全意识等指标。机械因素包括：各类机械设备的性能、可靠性、老化程度等指标。环境因素包括：自然环境及轨道交通系统环境。管理因素包括：规章制度、作业标准、安全教育程度等指标。由于地铁从建设到运营涉及多个学科领域，

该系统上任何微小组织面临自然失效、扰动和蓄意破坏，都有可能在整个系统中引起不可预料的连锁反应[8]，因此其应急安全管理应考虑多要素联动治理[9] [10]。

资产管理。地铁系统的资产通常指的是地铁运营公司具有物资形态且可量化的价值项，包括线性资产、固定安装资产、可移动资产等[11]，具有种类繁多、区域分布零散、管理周期长、资金比重大等特点[12]。一些学者基于“智慧地球”、“智慧交通”等概念的提出，关注于宏观的资产管理、建设“智慧地铁”系统[13]，与传统资产管理系统相比，具有自主感知、实时诊断、自主决策、精准执行、自我学习等优势；另一些学者聚焦于资产设备在实际运营场景中可能出现的问题提出设施及指引标志的布局建议[14]。

网络化布局。地铁系统网络化基本特征是：运营4条及以上的多线路，实现中心城区1km半径内能找到站点且具有不少于3处的换乘站[15]。随着城市轨道交通的不断建设，地铁系统的布局和跨度也逐年递增，但目前国内城市地铁网络化运营水平不高[16]，网络化管理仍面临着诸多压力与挑战，也为风险防控能力及服务水平提升带来了更高的要求。毕小玉[17]研究了网络化运营背景下换乘站出现不同运营管理模式、不同清分方式时，AFC系统的设计及改造方案。袁也[18]等针对网络化背景下，乘客由于不同线路的运营时间差异而无法换乘的问题，以末班车换乘失败最小化为目标，将末班车衔接时刻表转化为可以运行在量子计算机上的二次无约束二值化优化问题。谢正光[19]等认为在当前新时代背景下，我国未来地铁交通管理的网络化不仅要“多网融合”，与其他交通方式协同发展，更要推进城市(群)一体化交通网，推动区域化网络布局。

能耗需求。绿色发展已成为当今各行各业运营管理的共性目标，一直以来，地铁的交通运输功能被看作低碳出行的代表，然而，近年来地铁系统能耗持续高走。针对此现象，张世成[20]等兼顾经济成本与低碳排放为双重目标，提出融合NSGA-II非支配排序与动态拥挤测率的改进粒子群优化算法，建立成本与碳排放的动态在线评估模型。李晓峰[21]等以全生命周期分析为指导，覆盖车站建造、建材生产、建材运输、车站运营、列车运行等全流程的降碳方案。周勇[22]等通过综合选线、优化敷设方式，创新应用高架线路综合减振降噪技术、预制装配式绿色建筑板材，结合智慧化云平台等系统性技术措施，探索系统性创新的协同低碳方案。

3. 随机森林——隐马尔可夫模型

3.1. 隐马尔可夫模型原理

隐马尔可夫模型(Hidden Markov Model, HMM)是马尔可夫链的一种，最早由L E Baum等人提出[23]，是一种关于时间序列的概率模型，被广泛应用于故障诊断[24]、图像处理[25]、轨迹预测[26]等领域。一个HMM模型包括以下五个关键要素：观测状态序列 O 、隐状态序列 Y 、初始状态概率 $\pi(H_i)$ 、观测概率 $B(H_i, O_h)$ 、状态转移概率 $A(H_i, H_j)$ 。隐马尔可夫作为一种经典的时序建模工具，利用其对地铁站乘客动线进行分析，主要具有以下几点优势：

时序依赖性。以出行为目的的乘客在地铁站内的通行路径选择并非随机、无序的行走，而是遵循一定的动线流程约束，HMM的马尔可夫性正好契合这种“流程驱动”的特性，通过建立状态转移矩阵，能够学习乘客呈现出不同隐状态时的转换规律。

适用于零散点位观测。地铁站内的乘客动线观测为零散的观测点，HMM能为部分可观测的马尔可夫过程提供序列建模与状态推断框架，推断出最为可能的完整状态序列。

概率推断框架的鲁棒性。相比基于确定性规则或启发式搜索的方法，HMM的概率推断框架对噪声数据和异常值具有更强的鲁棒性。模型通过融合序列全局信息进行决策，前向-后向算法能够计算每个时刻隐藏状态的后验概率分布，为状态推断提供了不确定性度量。这种基于全局优化的推断机制，有效减

少因局部匹配导致的误差累积问题。

潜在状态的建模能力。影响地铁站乘客的行为决策的偏好中，存在一些无法直接观测但会影响设施选择的因素，HMM 可以通过隐状态来描述这些潜在的行为偏好，将其与设施选择及通行速度建立关联，符合乘客路径选择的决策机制。

根据 HMM 模型五要素，结合本文目标及现有数据情况，可将各个要素对应到本文案例中详细解释为：

观测状态序列，将乘客在设施之间的转移视为观测状态。仅统计合规动线，对于违反通行规则的路径在训练集中转移概率应为 0。对于每一次设施节点间的转移 $\tau = \{(u,v)\}$ ，构建 5 维特征向量：

$$x'_{u,v} = \{start_{code}, end_{code}, duration, distance, speed\}$$

其中， $start_{code}$ 为起始设施节点编码； end_{code} 为终止设施节点编码； $duration$ 为相邻两设施节点之间的通行时间； $distance$ 为相邻两设施节点之间的通行距离； $speed$ 为相邻两设施节点间的通行速度，取该段的平均速度。

隐状态序列，定义为乘客的行为偏好或出行意愿，主要反映的是乘客在设施节点间转移时的习惯或偏好。由于视频监控覆盖范围有限且乘客出行意愿无法直接观测，可将其作为本文建模中的隐状态。按照《地铁安全疏散规范》中对地铁站内乘客水平方向移动速度的定义，将隐状态的标注进行标准化如图 1 所示，乘客常见的出行意愿包括以下三类，分别对应：

状态 1：通行效率优先。此类乘客在前方无障碍及拥堵的情况下，按照较快速度步行或奔跑通过站内通道及设施节点。场景常见于中青年乘客的日常通勤、中小學生上下学出行等。

状态 2：通行便利优先。此类乘客通常对该地铁站内设施布局并不熟悉或者由于结伴出行需要等待他人的进程。场景常见于接送上下学的家长及儿童、共同旅游或出差的外地乘客等。

状态 3：通行稳定优先。此类乘客更注重出行过程的平稳性而非通行效率，在前方无障碍及无拥堵的情况下，始终保持正常或较慢速度匀速步行。场景常见于非高峰时段的日常出行人群、携带少量随身物品的旅客以及无明确时间紧迫性的休闲出行者等。

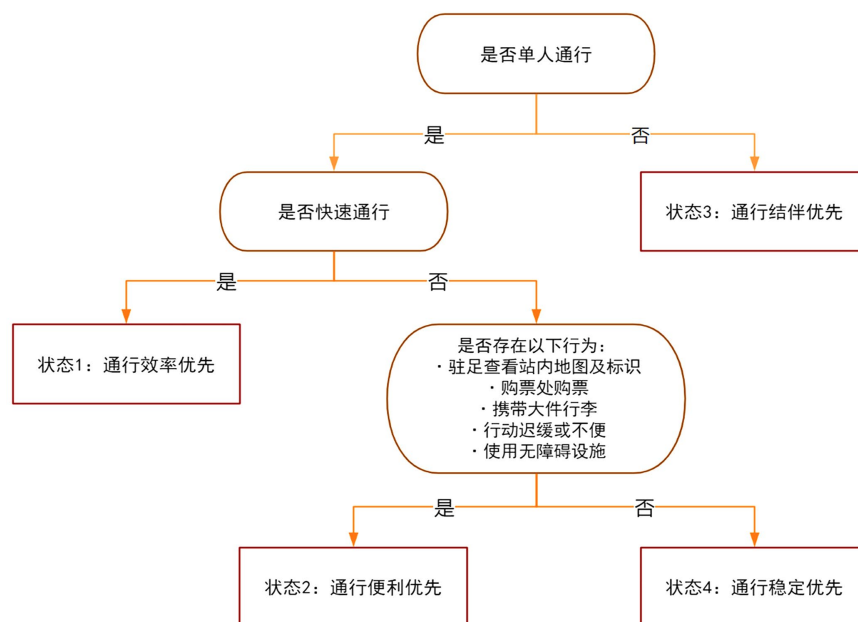


Figure 1. Standardized definition of hidden state determination
图 1. 隐状态判断标准化定义

初始状态概率，初始状态指的是乘客第一条有效转移 $f_{p,1} \rightarrow f_{p,2}$ 对应的隐状态， $\pi(H_i)$ 表示初始隐状态为 H_i 的概率，采用频率估计法计算，考虑到训练集样本较少，且在监控时段内存在有些进站通道无乘客进入，在此引入拉普拉斯平滑，以避免初始概率为 0 的情况：

$$\pi(H_i) = \frac{N_{init}(H_i) + \alpha}{\sum_{i=1}^3 N_{init}(H_i) + n \cdot \alpha} \quad (1)$$

其中， $N_{init}(H_i)$ 为训练集中初始隐状态为 H_i 的乘客数； $\sum_{i=1}^3 N_{init}(H_i)$ 为初始样本总数； $\alpha = 10^{-5}$ ，平滑系数； $n = 3$ ，隐状态数量。

观测概率，观测概率 $B(H_i, O_t)$ 表示隐状态为 H_i 时，生成当前观测特征向量 O_t 的似然概率，利用训练集中的乘客通行数据，计算得到先验概率：

$$P_{prior}(H_i) = \frac{N_i^{total} + \alpha}{\sum_{j=1}^3 N_j^{total} + n \cdot \alpha} \quad (2)$$

其中， N_i^{total} 为训练集中 H_i 出现的总次数。

传统的 HMM 假设观察概率服从高斯分布，但地铁通行场景下乘客通行特征包含离散变量和非线性关系。为保证特征间的相关性，引入随机森林分类器代替参数化的观测概率分布，使得模型能够更好的捕捉乘客的通行意志偏好与设施选择之间的复杂关系。

首先，利用预训练的随机森林分类器作为发射模型，输入当前时刻的特征向量 O_t ，输出该特征属于各状态的后验概率 $P_{RF}(H_i | O_t)$ ，引入贝叶斯矫正将随机森林输出的后验概率转化为发射概率：

$$B_{H_i} = \frac{P_{RF}(H_i | O_t)}{P_{prior}(H_i)} \quad (3)$$

其中， $P_{prior}(H_i)$ 为训练集中隐状态 H_i 的先验概率。

状态转移概率，描述隐状态之间按时间顺序转移的概率矩阵，即乘客通过某一设施节点后，从一种状态转移到一种状态的概率。

$$A(H_i, H_j) = \frac{C(H_i, H_j) + \alpha}{\sum_{i=1}^3 \sum_{j=1}^3 C(H_i, H_j) + n \cdot \alpha}, i, j \in (1, 2, 3) \quad (4)$$

其中， $C(H_i, H_j)$ 表示乘客在当前区间隐状态为 H_i 时、后一转移区间隐状态为 H_j 的合规连续转移数量，反映了乘客行为偏好的动态变化。

3.2. 核心集合定义

设施集合： $F = \{I_1, \dots, I_9, C_1, \dots, C_4, EN_{11}, \dots, EN_{33}, RE_{01}, S_1, \dots, S_{10}\}$ ，表示地铁站内所有的设施节点。

隐状态集合： $Y = \{H_1, H_2, H_3\}$ ，分别对应“通行效率优先”、“通行便利优先”、“通行稳定优先”。

设施节点间转移： $\tau = \{(u, v)\}$ ，定义为乘客从设施 u 到设施 v 的单次转移，其中 $u, v \in F$ 且满足合规性要求称为“可达 (u, v) ”，每个合规区间单一对应一个隐状态。

观测状态集合： $O = \{f, t\}$ ，从监控数据中可以直接获取到一个观测向量的 2 维特征：设施编码 f 及设施间的通行时间 t 。

乘客轨迹集合： $Tr = \{tr_1, tr_2, \dots, tr_N\}$ ， N 为训练集乘客数量，单名乘客 p 的轨迹

$tr_p = \{(f_{p,1}, t_{p,1}, H_{p,1}), (f_{p,2}, t_{p,2}, H_{p,2}), \dots, (f_{p,K}) | f_{p,1}, \dots, f_{p,K} \in F\}$ ，其中 K 为乘客 p 到达的设施节点数， $t_{p,K-1}$ 为乘客 p 从设施 $f_{p,K-1}$ 通行至 $f_{p,K}$ 的通行时间， $H_{p,K-1}$ 为乘客 p 从设施 $f_{p,K-1}$ 通行至 $f_{p,K}$ 的隐状态，

由于最后一个设施 $f_{p,K}$ 为终点设施，因此无转移至下一设施节点的时间和隐状态。

3.3. 乘客轨迹合规性约束

为确保乘客通行数据有效且符合地铁站内实际情况，乘客合规轨迹 $Tr^* \subseteq Tr$ 需满足以下约束条件：

安检规则：由于进站口地形限制，存在不同进站通道共用一个安检设施的情况，即当 $f_{p,k} \in \{I_1, I_5, I_9\}$ 时， $f_{p,k+1} = C_1$ ； $f_{p,k} \in \{I_2, I_6\}$ 时， $f_{p,k+1} = C_2$ ； $f_{p,k} \in \{I_3, I_7\}$ 时， $f_{p,k+1} = C_3$ ； $f_{p,k} \in \{I_4, I_8\}$ 时， $f_{p,k+1} = C_4$ 。

可达性规则：设施间距离矩阵 $D(u,v) = d$ ，当 $u \rightarrow v$ 可达时， $D(u,v) = d$ 表示为设施 u 到 v 的通行距离为 d ；当不可达或不符合通行规范时， $D(u,v) = 0$ 。乘客 p 的通行轨迹应满足 $D(f_{p,K}, f_{p,K+1}) \neq 0$ ，对于不符合可达性规律的片段可直接剔除。

时间单调性：对于任意的 $1 \leq k \leq K$ ，有 $t_{p,K} < t_{p,K+1}$ 。

4. 乘客轨迹还原

Viterbi 算法作为 HMM 的核心工具，用于解决观测序列与隐状态序列之间的解码，目标是找到观测序列下最为可能的隐状态，使得联合概率最大。

起始阶段：

定义向前变量 $\delta_1(i)$ 为到达时刻 1、状态为 H_i 时的最大路径概率：

$$\delta_1(i) = \pi(H_i) \cdot P_{prior}(H_i) \quad (5)$$

前一时刻的状态指引，可以路径回溯指针表示为：

$$\varphi_1(i) = 0$$

中间阶段：

对于 t 时刻、状态为 H_j 时的最大路径概率：

$$\delta_t(j) = \max_{i \in \{1,2,3\}} [\delta_{t-1}(i) \cdot a_{ij}] \quad (6)$$

当有最大值时，对应的路径回溯指针为：

$$\varphi_t(j) = \arg \max_i [\delta_{t-1}(i) \cdot a_{ij}] \quad (7)$$

终止阶段：

确定最优路径的终点状态：

$$h_T^* = \arg \max_i \delta_T(i) \quad (8)$$

通过回溯指针恢复完整路径：

$$h_t^* = \varphi_{t+1}(h_{t+1}^*), t = T-1, T-2, \dots, 1 \quad (9)$$

为避免概率连乘导致的数值下溢，采用对数空间动态规划：

$$\log \delta_t(j) = \max_i [\log \varphi_{t-1}(i) + \log A(H_i, H_j)] + \log P_{prior}(H_j) \quad (10)$$

并在每一步进行归一化为：

$$\delta_t(j) \leftarrow \frac{\delta_t(j)}{\sum_{k=1}^3 \delta_t(k) + \epsilon} \quad (11)$$

其中， $\epsilon = 10^{-10}$ 为稳定常数。

为学习从特征向量 x_i 到隐状态 H_i 的映射, 采用基尼不纯度作为决策树节点分裂准则, 对于节点 m 的候选分裂, 基尼不纯度定义为:

$$G(m) = 1 - \sum_{i=1}^3 p_i^2 \quad (12)$$

其中, p_i 为节点 m 中属于状态 H_i 的样本比例。

最优分裂选择使加权基尼不纯度下降最大的特征和阈值:

$$\Delta G = G(m) - \frac{|D_{left}|}{|D_m|} G(D_{left}) - \frac{|D_{right}|}{|D_m|} G(D_{right}) \quad (13)$$

针对状态分布不平衡的问题, 使用类别加权策略, 将类别 H_i 的样本对基尼不纯度的贡献加权, 将 H_i 的权重定义为:

$$\omega_{H_i} = \frac{N}{K \cdot N_{H_i}} \quad (14)$$

其中, N 为总样本数; K 为隐状态类别数, 取值为 3; N_{H_i} 为 H_i 的样本数。

根据给定的零散观测点集合, 算法推断流程可分为三个阶段:

阶段一: 轨迹构建。利用拓扑约束和时间窗口匹配, 按照地铁站实际情况定义设施节点的合规后继关系, 将零散的设施节点观测数据组合为符合合规要求的完整动线:

$$succ(f) = \begin{cases} C_1, f \in I_1, I_5, I_9 \\ C_2, f \in I_2, I_6 \\ C_3, f \in I_3, I_7 \\ C_4, f \in I_4, I_8 \\ T_1, T_2, EN_{01}, \dots, EN_{33}, RE_{01}, f \in C_1, \dots, C_4 \\ EN_{01}, \dots, EN_{33}, RE_{01}, f \in T_1, T_2 \\ S_1, \dots, S_{10}, f \in EN_{01}, \dots, EN_{33}, RE_{01} \end{cases} \quad (15)$$

阶段二: 特征提取。按照建构的轨迹, 提取 5 维特征向量序列。

阶段三: Viterbi 解码。调用融入随机森林的 Viterbi 算法, 输出最优的隐状态序列。

5. 地铁站客流特征分析

5.1. 站内基本情况介绍

选取的研究对象为某二线城市交通枢纽换乘站, 内有 1 号线及 4 号线两条地铁线路交汇而过, 其中 1 号线途径两处火车站, 4 号线连通省立医院、会展中心等城市重要公共服务中心。该站平日客流量尚未饱和、无严重的站台拥挤, 节假日及高峰时段呈现客流量陡增[27], 客流特征复杂、人口流动性大、出行需求旺盛, 以该站点作为研究对象, 有利于理解乘客一空间的交互关系, 构建考虑多元出行特征下的乘客动线分析框架。

经现场实地观测及视频调查, 已知该站站厅层位于地下一层, 一号线南北走向、位于地下二层, 四号线东西走向、位于地下三层, 乘客进站有 A、B、C、D 四个进站口可选择, 每个进站口都设置有进出站扶梯及步梯各一处, A 口处设有无障碍电梯一部。四个进站口分别对应四台安检设备, 站厅层内南、北各设有售票机组, 站内设有进站闸机三组、出站闸机三组。乘客从站台层到达站厅层可选择步梯、扶梯或无障碍电梯, 各设施节点的编号及功能详见表 1。

Table 1. Facility node description
表 1. 设施节点说明

编号	节点名称	节点功能
I_1	A 口进出站步梯	乘客离开/进入地铁站系统的步梯通道
I_2	B 口进出站步梯	乘客离开/进入地铁站系统的步梯通道
I_3	C 口进出站步梯	乘客离开/进入地铁站系统的步梯通道
I_4	D 口进出站步梯	乘客离开/进入地铁站系统的步梯通道
I_5	A 口进出站扶梯	乘客离开/进入地铁站系统的扶梯通道
I_6	B 口进出站扶梯	乘客离开/进入地铁站系统的扶梯通道
I_7	C 口进出站扶梯	乘客离开/进入地铁站系统的扶梯通道
I_8	D 口进出站扶梯	乘客离开/进入地铁站系统的扶梯通道
I_9	A 口进出站无障碍电梯	乘客离开/进入地铁站系统的无障碍电梯
C_1	A 口安检设备	检查乘客及其携带物品是否符合进站安全要求
C_2	B 口安检设备	检查乘客及其携带物品是否符合进站安全要求
C_3	C 口安检设备	检查乘客及其携带物品是否符合进站安全要求
C_4	D 口安检设备	检查乘客及其携带物品是否符合进站安全要求
T_1	厅北售票机组	出售进站车票、提供线路查询
T_2	厅南售票机组	出售进站车票、提供线路查询
$EN_{01} - EN_{12}$	厅南进站闸机	乘客进站票务核验
$EN_{25} - EN_{33}$	厅北进站闸机	乘客进站票务核验
RE_{01}	厅西进站闸机	乘客进站票务核验
S_1	站厅北至一号线站台步梯	连接站厅与站台的乘客通行步梯
S_2	站厅南至一号线站台步梯	连接站厅与站台的乘客通行步梯
S_3	站厅东至四号线站台步梯	连接站厅与站台的乘客通行步梯
S_4	站厅西至四号线站台步梯	连接站厅与站台的乘客通行步梯
S_5	站厅北至一号线站台扶梯	连接站厅与站台的乘客通行扶梯
S_6	站厅南至一号线站台扶梯	连接站厅与站台的乘客通行扶梯
S_7	站厅东至四号线站台扶梯	连接站厅与站台的乘客通行扶梯
S_8	站厅西至四号线站台扶梯	连接站厅与站台的乘客通行扶梯
S_9	站厅至一号线站台 无障碍电梯	连接站厅与站台的乘客无障碍通行电梯
S_{10}	站厅至四号线站台 无障碍电梯	连接站厅与站台的乘客无障碍通行电梯

5.1. 地铁站数据预处理

本文的客流数据包括以下几部分：

数据一：训练集。提取自该地铁站 2025 年 2 月某工作日早高峰时期的进站客流监控视频，随机选取视频时段内的 114 名乘客作为记录样本，观测其在站内经过的所有设施节点编号及到达该节点对应的时间。

数据二：预测集。提取自该地铁站 2025 年 5 月某工作日早高峰时期的进站客流监控视频，在对乘客进站数据进行统计时将所有设施看作节点，记录乘客开始使用或通过的设施编号及对应的时间点，采集 36 分钟乘客到达不同设施节点处的时间作为客流样本信息。

数据三：测试集。提取自该地铁站 2025 年 5 月某工作日早高峰时期的进站客流监控视频(与数据二同一日)，以各个进站扶梯、楼梯、无障碍电梯为轨迹统计起点，随机选取视频时段内的 98 名乘客作为记录样本，观测其在站内经过的所有设施节点编号及到达该节点对应的时间。

为避免异常数据及无用数据对分析结果产生干扰，对数据分三步骤进行初步清洗：

简化数据结构。过滤掉数据中对特征分析无关的字段，仅保留乘客进、出闸机的站点名称及对应的时间点。

客流数据合并。地铁网络支持多种方式购票进站，乘客可根据自己的喜好及实际情况，选择持一卡通、手机端二维码支付、现场购买单程票、刷脸进站等多种进站方式，不同进站方式的数据存储信息不同、储存于不同的文件，需要将相同内容、不同进站方式的刷卡数据进行整理合并。

异常值剔除。对单边交易、短时同站进出等异常值进行剔除，仅保留运营时段内、符合进站乘车动线且动线完整的乘客轨迹数据。

5.2. 模型效果评价

基于以上 HMM-RF 模型，对现有数据情况进行训练并预测出更多的合规乘客动线，三种状态下的高频设施转移对如图 2 所示。

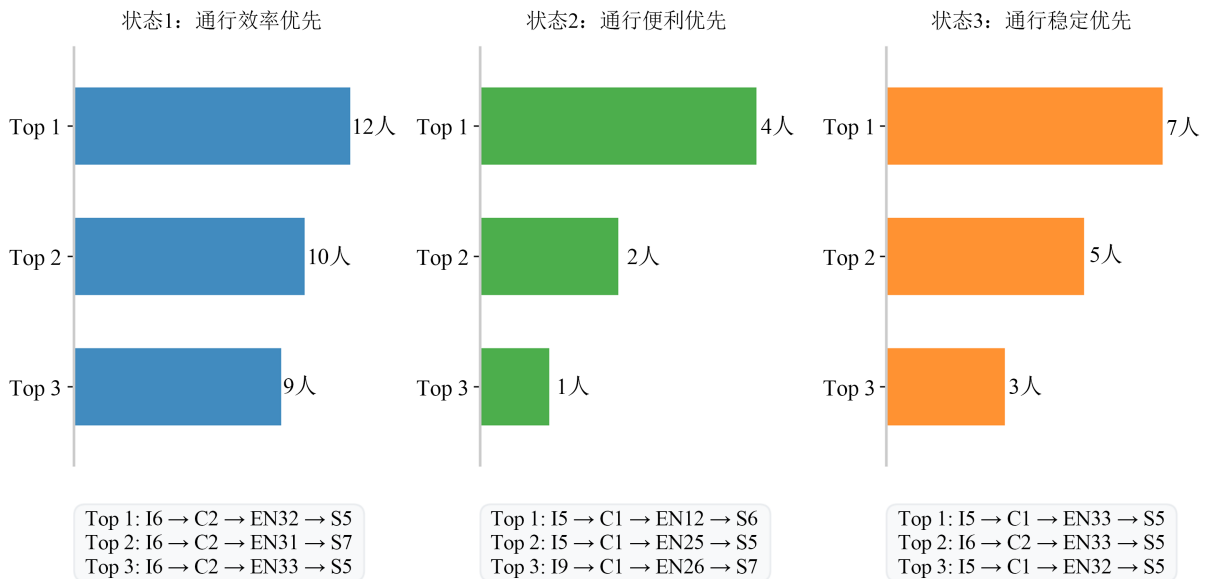


Figure 2. High-frequency facility transfer under three states

图 2. 三种状态下乘客高频选择设施转移路径

可以看出，不同出行意志的乘客对设施的选择明显不同，为体现公共交通服务的普适性，设施布局时除了考虑现有地形限制外，应充分考虑每类乘客的出行诉求。对于“通行效率优先”的乘客，为保证其核心诉求“快速通过”设施节点，一方面，应尽可能减少其高频动线与其他动线的交错，同时适当缩短设施之间的距离，减少通行路径上的干扰因素；另一方面，提升高频设施对中设施的使用快捷性，例如在闸机上配置“刷脸进站”等快速进站方式，通过提升设施通行的流畅度以缩短通行所需时间。对于

“通行便利优先”的乘客，应确保其动线的无障碍性，减少上、下行路段，尽可能保障通行路径的畅通、平坦；同时，可适当扩宽设施及等候区域的空间，避免大件行李对通行造成的拥挤。对于“通行稳定优先”的乘客，可优化其动线设施的舒适性，在设施转移节点处增加清晰的指示标识，特别是售票处、咨询处等服务节点；同时，考虑到动线上设施的舒适性，结合其停留需求，可增加休憩区提升出行体验。

按照前文提到的步骤对模型进行训练，训练完成后，使用验证集对最终模型进行性能评估打分，评估指标包括整体准确率(Accuracy)、整体精确率(Precision)、整体召回率(Recall)及 F1 分数。此外，绘制 3 类隐状态的 ROC 曲线，进一步从多维度分析模型对乘客在不同隐状态下的识别能力与泛化水平，模型综合评分结果如图 3 及图 4 所示。

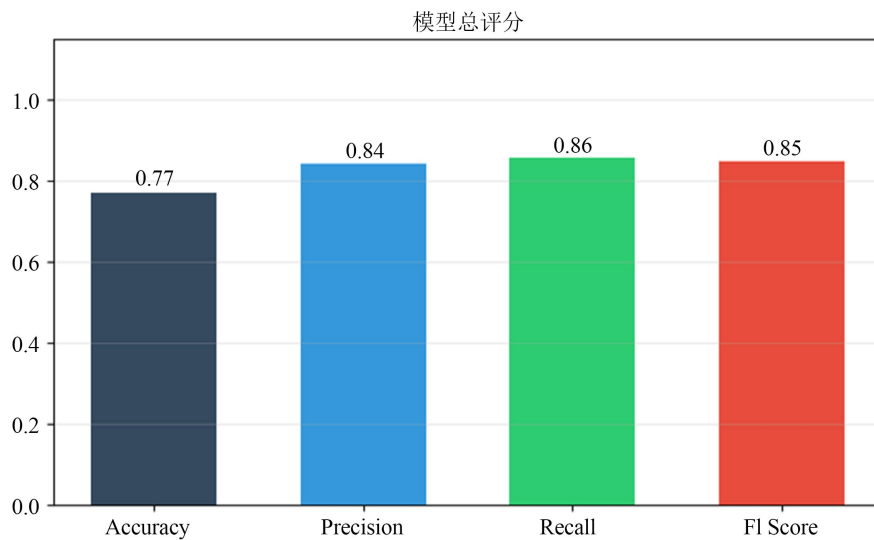


Figure 3. Comprehensive score of the HMM-RF model

图 3. HMM-RF 模型综合评分

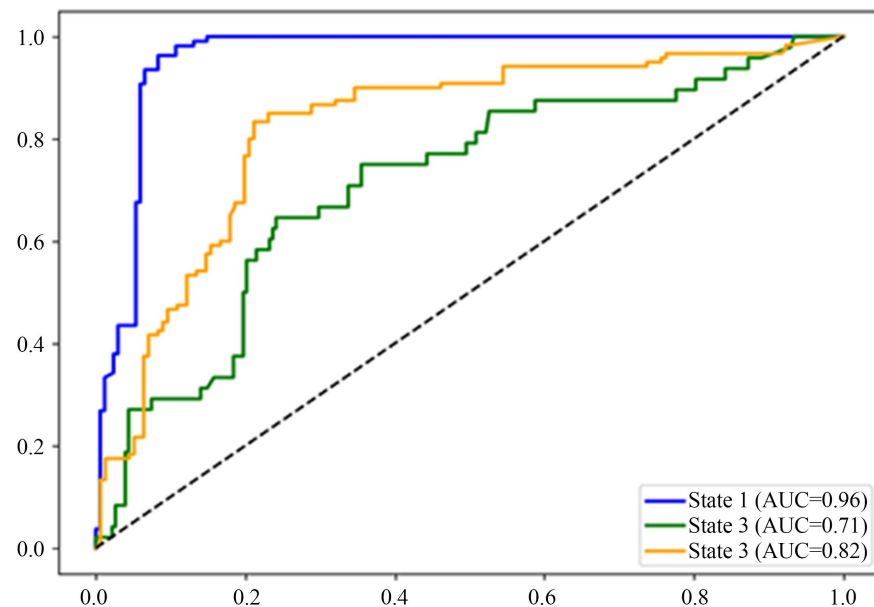


Figure 4. ROC curve for the HMM-RF model

图 4. HMM-RF 模型 ROC 曲线图

按照与上述融合模型相同的参数设置，另设置两组实验对照组：仅利用 HMM 模型和仅利用随机森林模型对相同的数据进行预测并对预测结果以相同的标准进行评分，三种不同模型的预测结果评分如表 2 所示。

Table 2. Performance comparison of different models
表 2. 不同模型性能对比

方法名称	Accuracy	Precision	Recall	F1 分数	AUC
HMM	0.46	0.46	0.46	0.45	0.59/0.53/0.54
Random Forest	0.44	0.41	0.44	0.32	0.59/0.65/0.50
HMM-RF	0.77	0.84	0.86	0.85	0.96/0.71/0.82

相较于纯 HMM 模型和纯随机森林模型，HMM-RF 混合模型在各个评分指标上都呈现出明显的提升效果，三个隐状态中状态 2——通行便利优先的 AUC 值相对较低，但也表现出较高的判别能力，评分超过 0.7、达到了可接受水平。总体而言，利用 HMM-RF 混合模型进行乘客轨迹还原实现了较好的动线分析效果，能有效减少地铁站乘客完整动线获取的时间和难度。

6. 结语

综上所述，本文所构建的 HMM-RF 混合模型，不仅在理论上解决了多源异构数据下的时空序列建模问题，更在工程实践上提供了一套将“碎片数据”还原为“完整链条”的处理流程，使原本模糊的乘客微观行为变得清晰可测，解析出的乘客动线数据将直接作为输入，为后续地铁站设施布局优化的方案设计提供有效依据。

参考文献

- [1] 刘力铭. 重庆市轨道交通公共服务质量提升路径研究[D]: [硕士学位论文]. 重庆: 重庆大学, 2024.
- [2] 李玉书, 孙越, 万衡, 等. 城市轨道交通车站换乘通道客流压力的评估方法[J]. 城市轨道交通研究, 2020, 23(1): 106-109, 144.
- [3] 刘成圻. 重庆轨道交通服务质量评价与提升策略研究[D]: [硕士学位论文]. 重庆: 重庆工商大学, 2025.
- [4] 刘瀚中, 刘志钢. 基于结构方程模型-模糊综合评价法的地铁乘客满意度研究[J]. 物流工程与管理, 2023, 45(3): 101-105, 96.
- [5] 于瀛. 轨道交通站内智能储物设施研究与设计[J]. 机械管理开发, 2023, 38(7): 91-93, 96.
- [6] 肖逸影, 王东, 吴宁宁, 等. 冻雨极端天气下武汉市轨道交通客流特征分析[J]. 现代城市轨道交通, 2025(7): 107-113.
- [7] 高嘉豪, 原培, 肖赵文, 等. 城市轨道交通运营安全评价研究[J]. 黑龙江科学, 2025, 16(17): 159-161.
- [8] 夏泽郁, 汤育春, 李启明. 基于韧性理论的中国城市轨道交通事故统计分析[J]. 都市快轨交通, 2020, 33(3): 148-156.
- [9] 孙沈强. 城市轨道交通运营过程中突发事件的安全管理对策[J]. 人民公交, 2025(16): 171-173.
- [10] 杨晓霞, 张蕊, 李永行, 等. 火灾爆发时地铁站乘客疏散多目标路径优化方法[J]. 交通运输工程学报, 2023, 23(5): 192-209.
- [11] 赵刚. 城市轨道交通资产管理信息系统建设研究[J]. 城市轨道交通研究, 2022, 25(4): 6-9.
- [12] 江婷婷, 安雪晖. 资产管理系统在国内地铁运营商中的应用[J]. 清华大学学报(自然科学版), 2015, 55(2): 176-183.
- [13] 赵刚. 基于“智慧地铁”的城市轨道交通资产管理信息系统研究[J]. 城市轨道交通研究, 2022, 25(1): 166-169, 175.
- [14] 刘枫, 邵春福, 贾洪飞, 等. 城轨换乘站服务设施布局与流线优化仿真评价[J]. 系统仿真学报, 2021, 33(7): 1670-1681.
- [15] 黄攀. 城市轨道交通网络化运营带来的问题及应对措施[J]. 城市轨道交通研究, 2023, 26(6): 247-250.

-
- [16] 雷强. 适应网络化运营的新建城市轨道交通线路运营组织方案设计研究[J]. 现代城市轨道交通, 2024(7): 115-118.
- [17] 毕小玉. 网络化运营背景下城市轨道交通换乘站 AFC 系统改造方案研究[J]. 城市轨道交通研究, 2022(S2): 75-78.
- [18] 袁也, 徐皓, 王悉, 等. 基于量子计算的城市轨道交通网络末班车衔接优化[J]. 都市快轨交通, 2024, 37(2): 139-145.
- [19] 谢正光, 魏运. 新时代我国城市轨道交通运营新模式探讨[J]. 都市快轨交通, 2022, 35(1): 54-59.
- [20] 张世成, 李伟, 蒲浩, 等. 地铁线路纵断面经济-低碳双目标智能优化[J]. 铁道科学与工程学报, 2026, 23(1): 124-136.
- [21] 李晓锋, 王荔, 栾承志, 等. 基于实际地铁线路的全生命周期碳排放研究[J]. 都市快轨交通, 2024, 37(1): 82-87.
- [22] 周勇, 彭帅, 周超. 绿色低碳城市轨道交通规划设计与实践[J]. 城市轨道交通研究, 2025, 28(8): 7-13, 19.
- [23] Holgado, P., Villagra, V.A. and Vazquez, L. (2020) Real-Time Multistep Attack Prediction Based on Hidden Markov Models. *IEEE Transactions on Dependable and Secure Computing*, **17**, 134-147.
<https://doi.org/10.1109/tdsc.2017.2751478>
- [24] 郝芳, 王宏超, 李宏伟. 基于连续隐马尔可夫的滚动轴承故障诊断[J]. 中国工程机械学报, 2019, 17(2): 184-188.
- [25] Liu, X., Kuang, R.X., Wang, B. and Yue, Y.T. (2018) Segmentation Algorithm of Spine CT Image Based on Hidden Markov Random Field. *Journal of Harbin University of Science and Technology*, **2**, 1-5.
https://en.cnki.com.cn/Article_en/CJFDTOTAL-HLGX201802001.htm
- [26] 陶希祥. 建筑垃圾运输车辆轨迹地图匹配与偏离判断研究[D]: [硕士学位论文]. 北京: 北京交通大学, 2022.
- [27] 林晓锋, 邹志光. 基于城市类型的我国城市轨道交通制式建设现状分析[J]. 现代城市轨道交通, 2025(2): 11-17.