Computer Science and Application 计算机科学与应用, 2026, 16(11), 76-88
Published Online November 2026 in Hans. https://www.hanspub.org/journal/csa
https://doi.org/10.12677/csa.2026.1611327

Hans汉斯

基于多源数据融合与GMM时空基线的车辆异常轨迹检测与风险预警平台设计与实现

侯学慧*，冯玉花

新疆警察学院信息安全工程系，新疆 乌鲁木齐

收稿日期：2026年10月4日；录用日期：2026年11月3日；发布日期：2026年11月11日

摘　要

针对车辆轨迹数据来源异构、时空连续性不足以及固定规则难以刻画不同车辆正常出行规律的问题，本文设计并实现一套基于多源数据融合与个性化时空基线的车辆异常轨迹检测与风险预警平台。平台以车辆GPS定位、道路卡口识别以及停车场出入口记录等多源轨迹数据为基础，通过坐标标准化、时间对齐、插值补全和卡尔曼滤波完成异构轨迹融合；进一步提取停留区域、常驻区域和高频路线等车辆出行特征，并将空间位置与小时、星期周期特征联合建模，采用高斯混合模型(GMM)建立车辆历史正常出行时空基线。在此基础上，以轨迹点在基线模型中的负对数似然作为异常分数，依据历史正常样本异常分数分布设置阈值，并将异常程度映射为低、中、高风险及P3、P2、P1三级预警。平台采用Streamlit构建Web交互界面，结合Folium及地图瓦片服务实现轨迹地图、异常点、热力图、风险等级分布和预警记录的可视化展示。考虑真实车辆轨迹具有位置敏感性和业务数据安全要求，本文系统验证全部采用程序生成的虚拟多源轨迹数据，模拟结果仅用于验证“数据生成－轨迹融合－基线建模－异常检测－风险预警－可视化展示”的端到端流程，不用于表征真实业务场景下的检测精度。系统运行表明，所设计平台能够完成从多源轨迹处理到分级预警输出的完整功能闭环，可为车辆异常轨迹分析、风险研判及相关教学科研和原型系统构建提供工程化参考。

关键词

车辆轨迹，异常轨迹检测，多源数据融合，高斯混合模型，风险预警，数据可视化

Design and Implementation of a Vehicle Abnormal-Trajectory Detection and Risk-Warning Platform Based on Multi-Source Data Fusion and a GMM Spatiotemporal Baseline

*通讯作者。

文章引用: 侯学慧, 冯玉花. 基于多源数据融合与 GMM 时空基线的车辆异常轨迹检测与风险预警平台设计与实现[J]. 计算机科学与应用, 2026, 16(11): 76-88. DOI: 10.12677/csa.2026.1611327

Xuehui Hou*, Yuhua Feng

Department of Information Security Engineering, Xinjiang Police College, Urumqi Xinjiang

Received: October 4, 2026; accepted: November 3, 2026; published: November 11, 2026

Abstract

To address the heterogeneity of vehicle trajectory data sources, insufficient spatiotemporal continuity, and the limited ability of fixed-rule methods to characterize the normal travel patterns of different vehicles, this paper designs and implements a vehicle anomalous trajectory detection and risk early-warning platform based on multi-source data fusion and personalized spatiotemporal baselines. The platform integrates multi-source trajectory data, including vehicle GPS positioning data, road checkpoint recognition records, and parking-lot entrance and exit records. Coordinate standardization, temporal alignment, interpolation, and Kalman filtering are applied to fuse heterogeneous trajectory data and construct continuous trajectories. Vehicle mobility features, such as stay regions, frequently visited regions, and high-frequency routes, are then extracted. Spatial locations are jointly modeled with cyclic hour-of-day and day-of-week features, and a Gaussian Mixture Model (GMM) is employed to establish a personalized spatiotemporal baseline of historical normal travel behavior. The negative log-likelihood of each trajectory point under the baseline model is used as the anomaly score, and the anomaly threshold is determined according to the distribution of anomaly scores derived from historical normal samples. The degree of anomaly is further mapped to low-, medium-, and high-risk levels and to P3, P2, and P1 warning levels. A Streamlit-based Web interface, together with Folium and map tile services, is used to visualize vehicle trajectories, anomalous points, heat maps, risk-level distributions, and warning records. Considering the sensitivity of real vehicle location data and the security requirements of operational data, all system validation in this study is conducted using programmatically generated synthetic multi-source trajectory data. The simulation results are intended only to verify the end-to-end workflow of data generation, trajectory fusion, baseline modeling, anomaly detection, risk early warning, and visualization, rather than to represent detection accuracy in real-world operational scenarios. The system results demonstrate that the proposed platform can implement a complete functional workflow from multi-source trajectory processing to hierarchical warning output, providing an engineering reference for vehicle anomalous trajectory analysis, risk assessment, related teaching and research, and prototype system development.

Keywords

Vehicle Trajectory, Abnormal Trajectory Detection, Multi-Source Data Fusion, Gaussian Mixture Model, Risk Warning, Data Visualization

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

近年来，针对车辆轨迹异常检测问题，研究者开始从车辆运行过程中的多层次时空依赖关系出发，对异常轨迹进行建模与识别[1]-[3]。随着车辆定位终端、道路卡口、停车场与住宅区出入口等感知设备逐步形成连续的数据记录链路，车辆运行轨迹已成为时空行为分析与风险研判的重要数据基础。实际应用

中，同一车辆的轨迹信息往往来自不同采样频率、不同记录方式和不同空间精度的数据源，单一来源数据易受到采样稀疏、定位噪声、记录缺失等因素影响，难以直接形成稳定、连续且可用于建模的轨迹序列。因此，如何将多源异构轨迹进行统一表达、连续化处理，并在此基础上识别相对于历史正常模式的异常偏离，是车辆轨迹分析平台需要解决的核心问题。

已有研究通过地图匹配与轨迹聚类提取车辆常见路径模式，并依据轨迹与正常路径之间的偏离程度进行异常识别[4]。同时，已有警务数据可视化与建模平台研究表明，将数据预处理、模型分析和交互式可视化整合到统一平台，可降低多源数据分析链路的使用门槛，并增强结果的可解释性和可复现性。本文延续此前“数据处理–分析建模–交互可视化展示”的工程化设计思路[5]，将研究对象聚焦到车辆时空轨迹，以多源轨迹融合和车辆级历史行为基线为主线，构建异常检测与分级风险预警流程。与依赖固定时间、固定区域或单一距离阈值的规则方法相比，本文方法首先利用车辆历史正常轨迹学习其常见空间位置和时间周期分布，再根据新轨迹点对历史模式的偏离程度计算异常分数，从而形成面向车辆个体历史行为的相对异常判定。

本文工作的主要内容包括：1) 设计面向车辆 GPS 定位、道路卡口及住宅区或停车场出入口记录的多源数据融合流程，形成事件级融合结果和连续轨迹级融合结果；2) 构建包含空间位置、小时周期和星期周期特征的 GMM 时空基线模型，以负对数似然描述轨迹点对车辆历史模式的偏离程度；3) 建立“异常分数–风险等级-P1/P2/P3 预警”的分级风险映射机制；4) 基于 Streamlit 和 Folium 实现总览驾驶舱、轨迹地图、建模分析、预警中心和数据明细等平台功能，形成从数据处理到预警展示的闭环。需要说明的是，本文数据均由程序模拟生成，研究重点是方法流程和平台实现，而不是对真实车辆环境中的模型精度进行统计推断。

2. 需求分析与总体设计

2.1. 业务需求与功能划分

车辆异常轨迹分析需要同时覆盖轨迹接入、轨迹融合、规律建模、异常识别、风险解释和可视化呈现等环节。为保证分析链路完整，平台划分为数据接入、多源融合、特征分析、基线建模、异常检测与风险预警、可视化展示六类功能模块。数据接入模块负责形成不同来源的轨迹事件；融合模块负责统一坐标字段与时间轴；特征分析模块从连续轨迹中提取停留区域、常驻区域和高频路线；基线建模模块学习车辆历史正常时空分布；异常检测与预警模块输出异常分数和风险等级；可视化模块负责地图、图表、表格和数据导出。

2.2. 平台总体架构

平台采用“算法业务模块 + 可视化展示平台”的轻量化分层架构。算法层负责模拟数据生成、轨迹融合、特征提取、GMM 建模、异常检测和分级预警；展示层负责参数配置、结果装配、地图渲染、图表展示和数据导出。算法与展示解耦后，同一组结果对象可以在不同界面组件中复用，同时便于后续将模拟数据接口替换为真实业务数据接口(见图 1)。

2.3. 多源轨迹数据及虚拟数据说明

车辆轨迹重建通常需要综合不同来源的车辆时空数据，以提高轨迹信息的连续性和完整性[6]。系统采用 GPS 定位、道路卡口及停车场/住宅区出入口等多源车辆轨迹数据(见表 1)，并通过统一时间戳、车辆标识、经纬度、数据来源和事件类型等字段进行规范化处理，形成可用于轨迹融合与后续分析的统一事件结构。



Figure 1. Overall workflow of the vehicle trajectory anomaly detection and risk early warning platform
图 1. 车辆异常轨迹检测与风险预警平台总体流程

Table 1. Composition and roles of multi-source vehicle trajectory data
表 1. 多源车辆轨迹数据构成及作用

数据来源	车辆场景语义	主要字段/特征	在平台中的作用
GPS 定位	高频车辆位置上报	车辆 ID、时间、经纬度、速度	构成连续轨迹主体
道路卡口	车辆通过道路摄像头或卡口的识别记录	时间、卡口位置、车辆 ID、事件来源	补充关键道路节点与高可信事件
停车场/住宅区出入口	车辆进出受控区域的记录	时间、出入口位置、车辆 ID、进出类型	补充驻留起止与区域访问事件

考虑真实车辆轨迹可能包含敏感位置、重点区域访问规律和业务系统标识等信息，本文不直接接入真实车辆业务数据，而采用程序生成的虚拟多源轨迹数据进行原型验证。模拟数据按照预设的日常出行模式构造常驻位置、通勤路径和周期活动，并叠加随机定位噪声；实时测试阶段再在指定时段和指定区域注入偏离日常模式的异常轨迹。由此产生的异常点数量、异常比例和预警数量均受模拟规则及参数设置影响，仅用于验证系统流程是否能够正确执行和展示，不应解释为真实场景中的检测率、误报率或风险发生概率。

3. 车辆异常轨迹检测与风险预警方法

3.1. 多源异构轨迹数据融合

平台首先对不同来源的坐标字段进行标准化，将 GPS、卡口及出入口记录统一为经纬度表达。随后按时间合并各类事件，形成保留原始事件来源的事件级融合表。为满足后续连续轨迹分析需求，系统按照统一时间间隔对事件序列进行重采样，对缺失位置进行插值补全，并采用卡尔曼滤波降低随机

定位噪声，最终形成轨迹级融合结果。事件级结果用于追溯原始来源，轨迹级结果用于特征提取和模型训练。

在融合过程中，不同来源事件可配置不同置信度。设置 GPS、道路卡口和出入口记录分别采用 0.80、0.95 和 0.99 的基础置信度，用于体现结构化固定点事件相较于普通 GPS 位置点具有更高的空间确定性。该设定属于原型规则参数，真实部署时应根据设备精度、数据质量和业务系统可靠性重新标定。

3.2. 车辆行为特征提取

融合轨迹形成后，平台提取停留区域、常驻区域和高频路线等行为特征。停留点检测采用空间距离阈值和持续时间阈值联合判定：若车辆在给定半径内持续停留时间超过阈值，则将该连续片段聚合为停留事件。实现默认距离阈值为 80 m、持续时间阈值为 300 s。车辆场景中，这类停留区域可解释为常驻停车区域、单位或住宅区驻留点、临时长时停车点等。

常驻区域识别采用 DBSCAN 模型对融合轨迹进行空间聚类，以发现车辆长期反复出现的空间区域；高频路线则根据通勤时间段和轨迹点重复出现情况抽取常见移动路径[7]。

3.3. GMM 时空行为基线建模

高斯混合模型(Gaussian Mixture Model, GMM)是一种利用多个高斯分布加权组合描述复杂数据分布的概率模型，适合刻画车辆在多个常驻区域和不同时段下形成的多模态时空活动规律。

车辆的正常运行规律不仅表现为空间位置分布，也具有明显的时间周期性。若直接使用小时数和星期数作为线性特征，会造成 23:59 与 00:01 在数值上相距较远而在实际周期上相邻的问题。因此，平台将日周期和周周期映射为正余弦特征，并与轨迹经纬度共同构成 GMM 输入。设轨迹点 i 的时空特征向量为：

$$x_i = \left[lat_i, lon_i, \sin\left(2\pi t_i/24\right), \cos\left(2\pi t_i/24\right), \sin\left(2\pi d_i/7\right), \cos\left(2\pi d_i/7\right) \right] \tag{1}$$

其中，lat_i 和 lon_i 分别为轨迹点的纬度和经度，t_i 表示一天内的时间位置，d_i 表示星期位置。特征标准化后输入高斯混合模型。GMM 通过多个高斯分量刻画车辆在不同时间段和不同常驻区域的联合概率分布[8]，其概率密度可表示为：

$$p(x) = \sum_{k=1}^{K} \pi_k \cdot N\left(x \middle| \mu_k, \Sigma_k\right) \tag{2}$$

其中，K 为混合成分数，π_k 为第 k 个高斯分量权重，μ_k 和 Σ_k 分别为其均值向量与协方差矩阵。系统默认 GMM 成分数为 4，并根据样本规模限制最大有效成分数，以避免小样本情况下模型过度复杂。

对于 GMM 成分数 K，本文将 $K=4$ 作为原型系统的默认工程配置，不将其视为适用于所有数据分布的固定最优值。实际部署时可在候选 K 范围内利用贝叶斯信息准则(Bayesian Information Criterion, BIC)进行辅助筛选，并结合独立验证数据上的误报水平、检测稳定性和模型复杂度确定最终参数。

3.4. 异常分数与阈值判定

完成历史正常轨迹建模后，系统使用 GMM 对新轨迹点进行概率密度估计。若轨迹点出现在历史模型的高概率区域，则其行为与车辆既有规律较一致；若位于低概率区域，则表示其时间或空间位置偏离历史模式。本文采用轨迹点负对数似然作为异常分数：

$$S(x) = -\log p\left(x \middle| \theta\right) \tag{3}$$

其中，θ 表示 GMM 参数。异常分数越大，表示该轨迹点相对于历史正常行为基线的概率越低。平台以历史正常轨迹异常分数的 95%分位数作为基础异常阈值 T。当新轨迹点分数超过 T 时，将其视为异常候

选。该阈值属于统计型基线阈值，目的是使模型能够根据不同车辆自身历史分布进行相对判定，而不是使用固定空间半径或统一速度阈值。

异常阈值百分位同样属于需要校准的模型参数。原型系统采用 95%分位数以保持较高异常敏感性；在实际应用中，可根据独立验证数据、可接受的误报水平以及业务预警成本调整百分位，使异常检出与误报控制之间取得平衡。

3.5. 分级风险预警机制

为了将连续异常分数转换为更易理解的业务预警信息，平台在基础阈值 T 之上构建四级风险映射(见表 2)。规则为：R(x) = 正常，$S \le T$；低风险，$T < S \le 1.2\ T$；中风险，$1.2\ T < S \le 1.5\ T$；高风险，$S > 1.5\ T$。

风险等级进一步映射为 P3、P2 和 P1 预警。除异常分数外，系统还根据发生时段、与常驻区域的距离以及是否命中预设模拟异常区域生成触发原因，用于增强预警结果的解释性。系统还支持连续异常片段识别，当多个相邻轨迹点持续超过阈值时，可聚合为连续异常时段，避免仅从单点角度理解风险。

Table 2. Mapping of anomaly scores, risk levels, and warning levels
表 2. 异常分数、风险等级与预警级别映射

风险等级	分数区间	预警级别	原型处置建议
正常	$S \le T$	—	正常记录
低风险	$T < S \le 1.2\ T$	P3	记录日志，持续关注
中风险	$1.2\ T < S \le 1.5\ T$	P2	推送值班人员，人工核实
高风险	$S > 1.5\ T$	P1	立即上报，启动处置流程

本文风险等级主要依据轨迹异常分数进行规则映射，其目标是验证“异常检测–风险分级–预警输出”的完整处理链路，而不将异常分数直接解释为真实业务风险概率。实际应用中的风险等级还可能受到天气、节假日、区域功能属性、时段敏感性以及历史处置结果等上下文因素影响。因此，可在当前 GMM 异常检测结果基础上进一步构建独立的风险评估层，将异常分数与上下文特征联合输入逻辑回归、决策树等可解释模型，学习从异常状态到业务风险等级的映射。该方法需要具有可靠风险标签的历史数据进行训练和验证。由于本文采用程序生成的虚拟轨迹，缺乏来源于真实业务处置结果的风险等级标签，若直接依据人工规则生成标签并训练监督分类模型，其结果主要反映对预设规则的拟合能力，难以代表真实业务风险识别性能，因此本文暂不进一步训练监督式风险分类模型。

4. 平台实现与可视化集成

4.1. 算法模块与展示模块解耦

平台将核心算法与界面展示分离。算法模块统一输出 GPS 数据、卡口数据、出入口数据、事件级融合结果、轨迹级融合结果、停留区域、常驻区域、实时轨迹、异常检测明细、预警记录、风险统计及运行参数等对象；展示模块只负责读取这些结果并形成页面组件。该设计避免界面层再次生成独立数据，保证地图、图表、表格和指标卡片使用同一组算法输出。

4.2. Web 交互与地图可视化

平台前端采用 Streamlit 构建浏览器端交互界面，主要包含总览驾驶舱(图 2)、轨迹地图(图 3)、建模分析(图 4)、预警中心(图 5)、数据明细(图 6)和模块设置等页面。总览驾驶舱展示事件级融合点数、轨迹

级融合点数、实时轨迹点数、异常点数、预警条数、风险等级分布和异常分数时间轴；轨迹地图叠加历史轨迹、实时轨迹、异常点、常驻区域和热力图；建模分析页面展示停留区域、聚类结果和异常检测明细；预警中心支持按 P1/P2/P3 查看与导出预警记录。

地图显示层基于 Folium 和地图瓦片服务实现，同时提供 WGS84 到 GCJ-02、WGS84 到 BD-09 的坐标转换，以适配不同互联网地图底图。坐标转换仅发生于展示层，不改变算法层的原始轨迹坐标，从而将模型计算与地图显示逻辑分离。



Figure 2. Overview dashboard
图 2. 总览驾驶舱



Figure 3. Trajectory map
图 3. 轨迹地图



Figure 4. Modeling analysis
图 4. 建模分析



Figure 5. Warning center
图 5. 预警中心

4.3. 参数配置与结果追溯

平台允许配置历史学习天数、GPS 采样频率、GMM 成分数、异常发生时间、异常持续时间以及模拟异常区域位置等参数。所有原始多源数据、融合轨迹、异常检测结果和预警记录均可表格查看并导出 CSV，为模型调试、教学实验和结果复核提供数据基础。

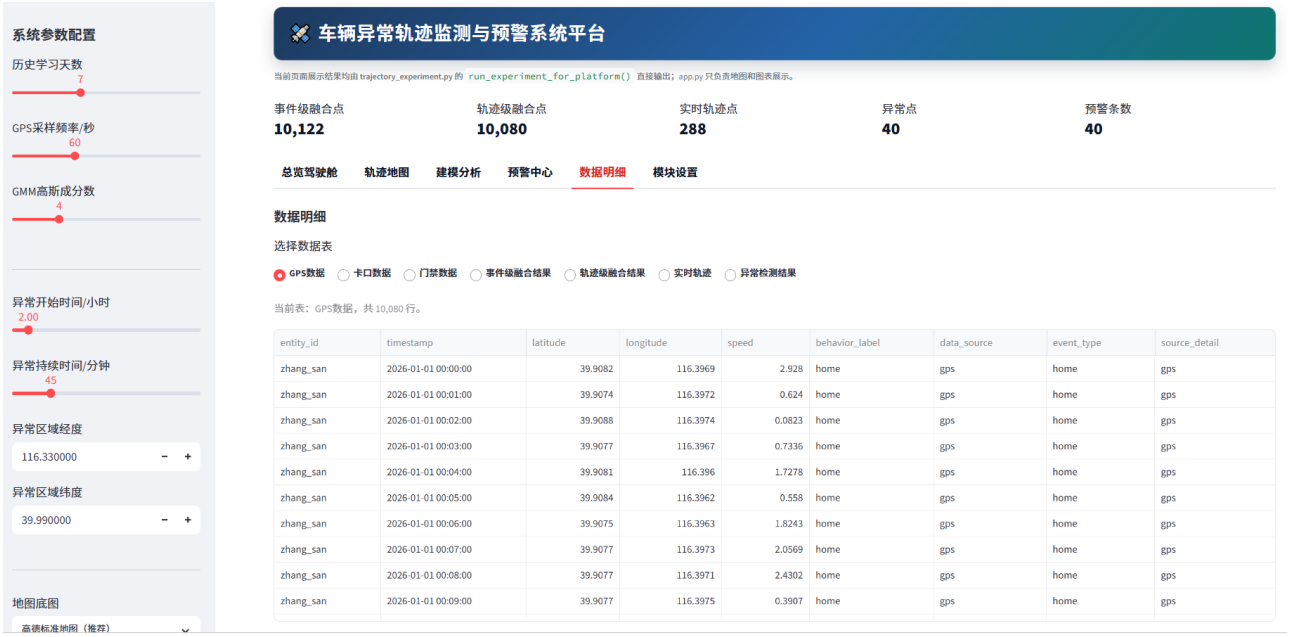


Figure 6. Data details
图 6. 数据明细

5. 系统验证与应用分析

5.1. 模拟场景与参数设置

已有研究利用生成式方法构造车辆轨迹并开展异常检测验证[9]。本文采用固定随机种子生成可重复的虚拟轨迹。实验主要参数见表 3，默认历史学习周期为 7 天，GPS 采样间隔为 60 s，GMM 成分数为 4；历史轨迹按照预设的日常驻留位置、通勤路线和周期活动生成，并叠加随机位置噪声。完成历史基线训练后，系统生成一段实时测试轨迹，并在凌晨 2:00 附近持续 45 min 注入远离日常常驻区域的模拟异常轨迹。实时轨迹按 5 min 间隔构造，用于观察异常分数、风险等级和预警记录的变化。

Table 3. Main experimental parameters
表 3. 实验主要参数

参数	默认值	用途说明
历史学习天数	7 天	构造车辆历史正常轨迹基线
GPS 采样频率	60 s	生成高频历史位置点
GMM 成分数	4	描述多模态时空活动分布
异常阈值	历史异常分数 95%分位数	形成相对异常判定基线
异常开始时间	02:00	构造凌晨异常场景
异常持续时间	45 min	形成连续异常片段
实时轨迹采样	5 min	生成测试轨迹

5.2. 异常检测流程验证

运行流程中，系统首先使用历史融合轨迹训练 GMM 并计算历史样本异常分数分布，再以 95%分位数形成阈值。随后将实时轨迹输入同一模型，逐点计算异常分数和风险等级。对于仍位于历史高概率区

域内的轨迹点，其分数通常处于阈值以内；当轨迹进入模拟的非日常区域后，空间位置与历史常驻模式显著偏离，同时异常发生在凌晨时段，因此异常分数上升并触发低、中或高风险标记。若连续多个点被判定为异常，系统将其识别为连续异常时段。

需要强调的是，上述现象来自预设规则与模拟轨迹共同作用。由于异常区域位置、持续时间和历史行为模式均由程序事先构造，系统在模拟案例中出现异常峰值是预期的功能性结果，其意义在于验证数据流、模型计算、阈值判定和风险分级能否按设计协同工作，而不能据此推导模型在真实道路环境中的召回率、精确率或泛化能力。

5.3. 分级预警与可视化结果

异常点被识别后，预警模块仅对异常记录生成预警表。预警信息包括预警时间、车辆标识、风险等级、预警级别、异常分数、触发原因、处置建议和位置坐标等字段。若轨迹点距离模拟异常区域中心小于 150 m，系统标记为“命中模拟异常区域”；距离小于 500 m 时标记为“接近模拟异常区域”。同时，若异常发生在 0:00~6:00，或距离历史常驻区域较远，则在触发原因中增加相应解释。

可视化层通过风险等级柱状图展示不同风险数量，通过异常分数时间轴呈现异常分数随时间变化，通过地图同时展示历史轨迹、实时轨迹、异常点、常驻区域和热力分布，并在预警中心展示 P1、P2、P3 数量及预警明细。由此形成从“异常发现”到“风险解释”再到“地图定位”的结果闭环。

5.4. 参数敏感性与消融实验

5.4.1. 实验设置与评价指标

为分析关键参数和不同类型特征对 GMM 异常检测结果的影响，在不改变主体数据生成与检测流程的前提下开展参数敏感性与特征消融实验。历史训练基线仍采用 7 天、60 s 采样的同一组融合轨迹；测试阶段保持原有“完整一天轨迹 + (02:00~02:45)远离常驻区域异常注入”的设置，并采用 20 组不同随机扰动的测试轨迹进行重复评价。每个测试点依据程序注入标记形成模拟真实标签，并逐点计算 Precision、Recall、F1、假阳性率(FPR)。需要说明的是，这些指标仅用于受控虚拟数据条件下比较参数和特征配置的相对变化，不代表真实道路环境中的实际检测精度。

5.4.2. 关键参数敏感性分析

Table 4. Sensitivity analysis results for the number of GMM components K (Threshold = 95%)
表 4. GMM 成分数 K 敏感性实验结果(阈值为 95%)

K	Precision	Recall	F1	FPR
2	0.3422	1.0000	0.5090	0.0629
3	0.3509	1.0000	0.5185	0.0606
4	0.2565	1.0000	0.4079	0.0943
5	0.1974	1.0000	0.3296	0.1317
6	0.1230	1.0000	0.2190	0.2315
7	0.1699	1.0000	0.2902	0.1584
8	0.7538	1.0000	0.8568	0.0111

首先考察 GMM 成分数 K 的影响。固定异常阈值规则为历史正常样本异常分数的 95%分位数，令 K 在 2~8 之间变化，其余训练数据、测试轨迹与模型设置保持一致。结果见表 4。当前模拟异常与正常轨迹

在异常分数排序上分离较明显，因此各 K 下 Recall 均为 1，差异主要体现在 Precision、F1 和 FPR。不同 K 的结果呈非单调变化，说明成分数会改变 GMM 对多模态时空分布的划分方式。本文仍保留 K = 4 作为原型默认配置，以保持与前述平台实验一致，但不将其解释为普适最优值；实际应用可结合 BIC 与独立验证指标进行自适应选择。

进一步固定 K = 4，将异常阈值百分位分别设置为 90%、92.5%、95%、97.5%和 99%。结果见表 5。随着阈值百分位提高，模型判定趋于保守，FPR 由 0.1815 持续下降至 0.0206，Precision 和 F1 相应提高；在当前强异常模拟条件下，各候选阈值的 Recall 均保持为 1。该结果表明阈值主要影响误报水平和预警输出规模。95%分位数能够保持较高异常敏感性，因此继续作为原型平台默认值；但在真实应用中应根据验证数据和业务可接受误报水平进行校准，而不应将 95%视为固定最优阈值。

Table 5. Sensitivity analysis results for the Anomaly-Threshold percentile (K = 4)
表 5. 异常阈值百分位敏感性实验结果(K = 4)

阈值百分位/%	Precision	Recall	F1	FPR
90	0.1516	1.0000	0.2631	0.1815
92.5	0.1858	1.0000	0.3132	0.1423
95	0.2565	1.0000	0.4079	0.0943
97.5	0.3791	1.0000	0.5488	0.0536
99	0.6135	1.0000	0.7595	0.0206

5.4.3. 特征消融实验

为量化空间特征和时间周期特征的贡献，在原始 GMM 模型上进行三组特征消融，除输入特征外其余条件全部保持一致：K 固定为 4，各模型均采用自身历史正常异常分数的 95%分位数作为阈值，并使用完全相同的训练轨迹和测试轨迹。仅空间模型使用经纬度二维特征，仅时间模型使用小时和星期的正余弦周期特征，时空联合模型使用原始 6 维特征，结果见表 6。

Table 6. GMM feature ablation results (K = 4, 95th-percentile threshold)
表 6. GMM 特征消融实验结果(K = 4，95%分位数阈值)

特征组合	Precision	Recall	F1	FPR
仅空间特征	0.2078	1.0000	0.3436	0.1244
仅时间特征	0.1800	1.0000	0.3051	0.1470
时空联合特征	0.2565	1.0000	0.4079	0.0943

三组模型在当前模拟强异常条件下均实现了 1 的 Recall，但时空联合模型的 Precision 为 0.2565、F1 为 0.4079、FPR 为 0.0943，均优于仅空间和仅时间配置。与仅空间模型相比，时空联合模型的 FPR 降低约 3.01 个百分点；与仅时间模型相比降低约 5.27 个百分点。说明空间位置是识别远距离轨迹偏离的重要基础，而时间周期特征与空间特征联合建模后能够进一步减少正常点误报，提高综合判别结果。由于本文测试异常具有较明显的空间偏离，三组模型的 Recall 均达到饱和，因此本节更关注 Precision、F1 和 FPR 的相对变化。

6. 讨论与改进方向

数据方面，当前平台采用程序生成的虚拟轨迹，优点是能够控制异常发生时间、位置和持续时长，

并避免真实车辆位置数据带来的隐私和业务安全风险，但模拟数据的行为复杂度明显低于真实道路环境。真实车辆可能受到交通拥堵、临时绕行、定位漂移、设备离线、驾驶人变化和节假日活动等多种因素影响。后续若具备合规数据条件，应建立匿名化车辆标识体系，在完成脱敏、访问控制、日志审计和网络隔离的前提下接入真实轨迹，并采用时间切分和跨车辆测试评价模型泛化能力。

方法方面，GMM 具有训练速度快、参数量小和结果可解释等优势，适合作为轻量化基线模型，但其主要刻画轨迹点的时空概率分布，对轨迹顺序、运动节奏和长距离路径结构的表达能力有限。后续可在保留 GMM 点异常能力的基础上，引入轨迹相似度、速度与加速度异常、LSTM 自编码器或 Transformer 时序模型，形成“点异常 + 序列异常 + 路径异常”的综合风险评分。

风险评估方面，当前 P1/P2/P3 分级采用基于异常阈值倍数的规则映射，适用于实验演示和流程验证。该机制具有实现简单和输出可解释的特点，但尚未充分利用天气、节假日、区域功能属性等外部上下文信息。后续在获得具有真实处置结果和风险等级标签的数据后，可将 GMM 输出的异常分数作为基础特征，并融合时段、区域属性、天气及历史核验结果等信息，采用逻辑回归、决策树等可解释监督模型学习风险等级映射，从而使风险评估由固定规则逐步转向数据驱动的自适应机制。

平台方面，当前实现支持单车辆分析和 CSV 文件格式的结果导出，后续可扩展多车辆批处理、轨迹相似度检索、实时流式数据接入、用户权限管理与闭环处置、自动报告生成等功能，并进一步增强多图联动和时间轴回放能力。

7. 结论

本文围绕车辆异常轨迹检测与风险预警需求，设计并实现了一套基于多源轨迹数据融合和 GMM 时空行为基线的可视化平台。平台将车辆 GPS、道路卡口及出入口记录统一到同一时间与空间表达，通过插值和卡尔曼滤波形成连续轨迹；在此基础上提取停留区域、常驻区域和高频路线等特征，并利用空间位置、小时周期和星期周期特征构建车辆历史正常行为 GMM 基线，以负对数似然计算轨迹异常分数，再依据历史分位数阈值和分数区间生成 P3、P2、P1 分级预警。Streamlit 与 Folium 实现了轨迹、异常分数、风险等级和预警记录的交互式可视化。

本文实验全部使用程序生成的虚拟多源轨迹数据，论文结论限定在方法流程和系统功能层面：模拟案例验证了“数据生成－融合－建模－检测－预警－展示”链路的完整性与可执行性，但不对真实车辆环境下的检测精度、误报水平或业务处置效果作出结论。后续研究将以合规、脱敏的真实数据为基础，围绕多车辆泛化、序列异常建模、阈值校准及业务闭环反馈开展进一步验证。

基金项目

新疆维吾尔自治区自然科学基金资助项目，项目编号：2025D01B94；2024 年度自治区高校本科教育教学研究和改革项目普通教改项目，项目编号：XJGXJGPTB-2024077；新疆警察学院校级科研创新团队及科研基地项目(数据建模与智能计算创新团队)，项目编号：XJKYKJTD202401。

参考文献

[1] 韩锋, 卜永丰, 梁浩翔, 等. 基于多层次时空交互依赖的车辆轨迹异常检测[J]. 计算机应用, 2026, 46(2): 604-612.

[2] Chen, J., Chen, H. and Lu, H. (2026) Enhancing Road Safety and Sustainability: A Multi-Scale Temporal Model for Vehicle Trajectory Anomaly Detection in Road Network Interactions. *Sustainability*, **18**, Article No. 597. https://doi.org/10.3390/su18020597

[3] Qiu, M., Mao, S., Zhu, J. and Yang, Y. (2025) Spatiotemporal Multi-Feature Fusion Vehicle Trajectory Anomaly Detection for Intelligent Transportation: An Improved Method Combining Autoencoders and Dynamic Bayesian Networks. *Accident Analysis & Prevention*, **211**, Article ID: 107911. https://doi.org/10.1016/j.aap.2024.107911

[4] 王书敏, 李元俊, 杨涛, 等. 基于地图匹配和两阶段聚类的轨迹异常检测算法[J]. 地球信息科学学报, 2025, 27(8): 1780-1795.

[5] 侯学慧, 冯玉花, 肖淑蓉. 面向警务业务的数据可视化处理平台设计与实现[J]. 计算机科学与应用, 2026, 16(3): 250-256.

[6] 郑慧敏, 许倩倩, 刘婉纯. 基于多源数据融合的城市主干道车辆轨迹重建技术研究[J]. 安阳工学院学报, 2026, 25(4): 88-94.

[7] 路强, 杨贵冰, 檀俊滔, 等. 道路交通趋势可视化中的多代表性轨迹聚类方法[J]. 计算机辅助设计与图形学学报, 2019, 31(7): 1194-1202.

[8] Wang, B., Li, W. and Khattak, Z.H. (2024) Anomaly Detection in Connected and Autonomous Vehicle Trajectories Using LSTM Autoencoder and Gaussian Mixture Model. *Electronics*, **13**, Article No. 1251. https://doi.org/10.3390/electronics13071251

[9] Shi, H., Dong, S., Wu, Y., Nie, Q., Zhou, Y. and Ran, B. (2025) Generative Adversarial Network for Car Following Trajectory Generation and Anomaly Detection. *Journal of Intelligent Transportation Systems*, **29**, 53-66. https://doi.org/10.1080/15472450.2023.2301691