



(12) 发明专利

(10) 授权公告号 CN 119380552 B

(45) 授权公告日 2025. 04. 04

(21) 申请号 202411898005.0
(22) 申请日 2024.12.23
(65) 同一申请的已公布的文献号
申请公布号 CN 119380552 A
(43) 申请公布日 2025.01.28
(73) 专利权人 交通运输部公路科学研究所
地址 100088 北京市海淀区西土城路8号
(72) 发明人 范青蓝 张一鹏 李振华 吴梦怡
刘砚玥 张卓敏 殷其昊
(74) 专利代理机构 成都创典汇智知识产权代理
事务所(普通合伙) 51383
专利代理师 张一帆
(51) Int. Cl.
G08G 1/01 (2006.01)
G08G 1/04 (2006.01)

G08G 1/048 (2006.01)
G06F 18/15 (2023.01)
G06F 18/214 (2023.01)
G06F 18/2431 (2023.01)
G06F 18/2433 (2023.01)
G06F 18/25 (2023.01)
G06N 3/045 (2023.01)
G06N 3/0464 (2023.01)
G06N 3/048 (2023.01)
G06N 3/0985 (2023.01)

(56) 对比文件

CN 118470974 A, 2024.08.09
CN 113743469 A, 2021.12.03

审查员 赵云峰

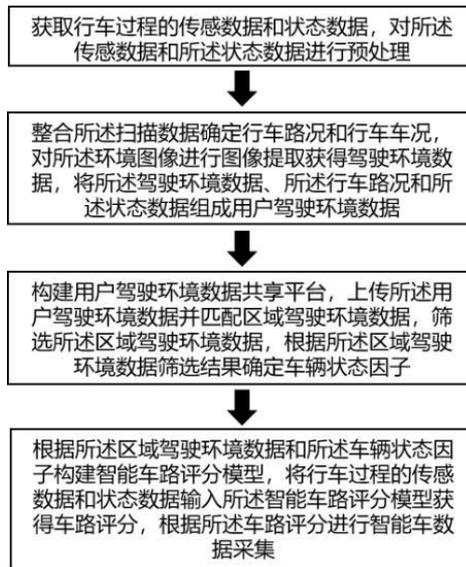
权利要求书4页 说明书10页 附图1页

(54) 发明名称

基于大数据人工智能分析的智能车数据采集方法及系统

(57) 摘要

本发明公开了基于大数据人工智能分析的智能车数据采集方法及系统,包括获取行车过程的传感数据和状态数据,整合所述扫描数据,对所述环境图像进行图像提取获得驾驶环境数据,构建用户驾驶环境数据共享平台并匹配区域驾驶环境数据,筛选所述区域驾驶环境数据,根据所述区域驾驶环境数据筛选结果确定车辆状态因子,根据所述区域驾驶环境数据和所述车辆状态因子构建智能车路评分模型,将行车过程的传感数据和状态数据输入所述智能车路评分模型获得车路评分,根据所述车路评分进行智能车数据采集。该方法不仅可以智能车数据采集的效率和准确性,同时具有较好的可解释性,可以直接应用于智能车数据采集系统中。



1. 基于大数据人工智能分析的智能车数据采集方法,其特征在于,包括以下步骤:

S1、获取行车过程的传感数据和状态数据,对所述传感数据和所述状态数据进行预处理;所述传感数据包括环境图像和扫描数据;

S2、整合所述扫描数据确定行车路况和行车车况,对所述环境图像进行图像提取获得驾驶环境数据,将所述驾驶环境数据、所述行车路况和所述状态数据组成用户驾驶环境数据;

S3、构建用户驾驶环境数据共享平台,上传所述用户驾驶环境数据并匹配区域驾驶环境数据,筛选所述区域驾驶环境数据,根据所述区域驾驶环境数据筛选结果确定车辆状态因子;

S4、根据所述区域驾驶环境数据和所述车辆状态因子构建智能车路评分模型,将行车过程的传感数据和状态数据输入所述智能车路评分模型获得车路评分,根据所述车路评分进行智能车数据采集;

根据区域驾驶环境数据筛选结果确定对应时刻的状态数据,将状态数据输入状态影响函数中获得车辆状态因子,表达式为:

$$Affect = w_1 \left[(1 - \sin \sqrt{U/M})(1 - \sin \sqrt{\delta})e^{-D} \right] \left\{ w_2 \left[v_t \ln \left(1 + \frac{v_t - v_{t0}}{v_{t0}} \right) + e^{-g_t} + e^{-\sin P_{pose}} \right] \right. \\ \left. + w_3 \sin \left(\sum_{i=1}^{n_1} c_i e^{P_i} + \sum_{j=1}^{n_2} c_j e^{\frac{H_j}{H_{j0}}} \right) + \frac{w_4}{T_w} \min (P_{pre}, P_{op}, P_{oq}) \right\}$$

其中 $Affect$ 为车辆状态因子, w_1 为失控概率权重, w_2 为速度状态因素权重, w_3 为安全驾驶因素权重, w_4 用户使用度权重, U 为辅助驾驶系统用户量, M 为额定用户量, δ 为辅助驾驶系统BUG概率, D 为网络延迟率, v_t 为当前车辆驾驶速度, v_{t0} 为当前时刻街道规定驾驶速度, g_t 为当前车辆加速度, P_{pose} 为当前车辆姿势评分, c_i 为车辆安全保护结构类别, n_1 为车辆安全保护结构类别数量, P_i 为车辆安全保护结构安全评分, c_j 为车辆驾驶影响因素类别, n_2 为车辆驾驶影响因素类别数量, H_j 为当前车辆驾驶影响因素值, H_{j0} 为当前车辆驾驶影响因素标准值, T_w 为当前车辆行驶时间, P_{pre} 为轮胎气压评分, P_{op} 为油箱油压评分, P_{oq} 为油箱油量评分;

将区域驾驶环境数据异常数据和对应车辆状态因子组成车路数据集,将车路数据集划分成训练集和测试集;

构建基于孤立森林的智能车路评分模型,输入层对输入数据进行预处理和特征选择,孤立森林模型为每个数据点输出异常评分,输出层根据异常评分和预设阈值为整个数据组分配一个整体评分、提取异常数据标签;

采用训练集训练孤立森林模型,采用交叉熵损失函数进行多类别异常划分,采用 RMSprop 优化器调整模型学习率,采用早停策略控制训练进程,采用测试集评估模型准确性;

将行车过程的传感数据和状态数据输入所述智能车路评分模型获得车路评分和对应的异常标签,根据车路评分选定智能车数据采集策略。

2. 根据权利要求1所述基于大数据人工智能分析的智能车数据采集方法,其特征在于,获取行车过程的传感数据和状态数据的方法,包括:

获得传感数据的方法包括:通过车身四周摄像头采集行车过程的环境图像,通过车身各处的激光雷达、毫米波雷达和超声波雷达采集行车过程的距离扫描数据;

获得状态数据的方法包括:通过GPS传感器采集行车定位数据,通过IMU传感器采集行车运动状态和姿态数据,直接通过车辆系统获取车辆状态数据。

3. 根据权利要求1所述基于大数据人工智能分析的智能车数据采集方法,其特征在于,整合所述扫描数据确定行车路况和行车车况的方法,包括:

将扫描数据划分为训练集和测试集;

构建行车数据模型,所述行车数据模型包括决策树、特征融合层和堆叠分类器;

决策树对输入扫描数据进行特征选择,并将扫描数据划分成行车路况和行车车况两类;

特征融合层包括多头注意力机制和特征融合策略,采用多头自注意力机制捕捉不同角度扫描数据间的空间关系、并行处理多个不同类别距离数据,采用Concatenate策略拼接融合不同维度的特征向量,表达式为:

$$Attention(Q, K, V) = \frac{\Phi(Q) \sum_{j=1}^n \Phi(K_j)^T V_j}{\Phi(Q) \sum_{j=1}^n \Phi(K_j)^T \sqrt{d_k}} + \lambda \cdot R(\Phi(Q), \Phi(K), \Phi(V))$$

$$MultiHead(Q, K, V) = Concat(head_1, head_2, \dots, head_h)W^0$$

$$head_i = Attention(QW_i^Q, KW_i^K, VW_i^V)$$

其中 $Attention(Q, K, V)$ 为注意力机制中查询向量 Q 在键值对上的注意力输出, K 为关注内容键向量, V 为关注内容值向量, $\Phi(\cdot) = elu(\cdot) + 1$ 为激活函数, K_j 为第 j 个关注内容键向量, V_j 为第 j 个关注内容值向量, n 为键值对数量, d_k 为尺度化因子, λ 为正则化强度超参数, $R(\cdot)$ 为L2正则化项, $MultiHead(Q, K, V)$ 为多头注意力机制表示,用于连接多个注意力, W^0 为初始参数矩阵, W_i^Q 为第 i 头查询向量参数矩阵, W_i^K 为第 i 头关注内容键向量参数矩阵, W_i^V 为第 i 头关注内容值向量参数矩阵, $head_i$ 为第 i 头输出;

堆叠分类器处理特征融合层输出数据,包括基模型和元模型,基模型间独立学习距离数据进行距离预测,元模型对基模型的距离预测结果进行最终决策,输出行车路况和行车车况;所述基模型包括随机森林、K近邻和支持向量机;

采用交叉熵损失函数区分行车路况和行车车况数据,采用网格搜索对行车数据模型参数进行调优,采用测试集数据对模型进行评估;

将待整合扫描数据输入行车数据模型获得行车路况和行车车况。

4. 根据权利要求1所述基于大数据人工智能分析的智能车数据采集方法,其特征在于,匹配所述区域驾驶环境数据的方法,包括:

将驾驶环境数据、行车路况和GPS定位信息组成用户驾驶环境数据;

构建用户驾驶环境数据共享平台,用户驾驶环境数据共享平台通过用户数据共享合约进行区域用户间驾驶环境数据的实时交互,采用对称加密算法对用户数据进行加密后传输,采用非对称算法验证用户身份和授权用户读取数据;

根据用户驾驶环境数据匹配用户驾驶环境数据共享平台数据库,定义本用户驾驶环境

数据向量为 $A = (a_1, a_2, \dots, a_d)$, 定义数据库中区域其他用户驾驶环境数据向量为 $B_j = (b_{j1}, b_{j2}, \dots, b_{jd})$, 其中 d 为向量维度, 根据数据相似度确定区域驾驶环境数据, 表达式为:

$$Sim_j = \frac{\sum_{i=1}^d (a_i \cdot b_{ji})}{\sqrt{\sum_{i=1}^d (a_i)^2} \sqrt{\sum_{i=1}^d (b_{ji})^2}} \left[\alpha \left(1 - \frac{cov(A, B_j)}{\sigma(A)\sigma(B_j)} \right) + \frac{\beta}{d} \sqrt{\sum_{i=1}^d (a_i - b_{ji})^2} \right]$$

其中 Sim_j 为本用户驾驶环境数据向量 A 与区域其他用户驾驶环境数据向量 B_j 的综合相似度, α 为差异波动占比, β 为幅值差异占比, 且 $\alpha + \beta = 1$, $cov(A, B_j)$ 为向量 A 与向量 B_j 的协方差, $\sigma(A)$ 为向量 A 的标准差, $\sigma(B_j)$ 为向量 B_j 的标准差;

选取与本用户驾驶环境数据向量 A 综合相似度最高的 m 个区域其他用户驾驶环境数据向量 B_j , 计算区域驾驶环境数据向量 $C = [A + \sum_{j=1}^m B_j] / (1 + m)$, 根据区域驾驶环境数据向量 C 确定区域驾驶环境数据。

5. 根据权利要求1所述基于大数据人工智能分析的智能车数据采集方法, 其特征在于, 筛选所述区域驾驶环境数据的方法, 包括:

采用等体积超球体法筛选区域驾驶环境数据, 设超球体为 $Ball(r_c, s)$, r_c 为超球体半径, s 为超球体维度, $Vol[Ball(r_c, s)]$ 为超球体体积, 通过搜寻最小星偏差优化等效半径 r_c , 表达式为:

$$\hat{R}_{EVHS} = arg \left\{ \min_{\hat{r}_c \in r_c} \sup \left| \frac{N[Ball(\hat{r}_c, s), \hat{R}]}{o} - \frac{Vol[Ball(r_c, s)]}{Vol[Cube(2\vartheta, s)]} \right| \right\}$$

其中 \hat{R}_{EVHS} 为搜寻最小星偏差得到的均匀性最好的有效代表点集, \hat{R} 为参考点集, \hat{r}_c 为最优等效半径, $N[Ball(\hat{r}_c, s), \hat{R}]$ 为半径为 \hat{r}_c 的超球体 $Ball(\hat{r}_c, s)$ 内有效代表点数量, o 为总的的数据点数量, $Vol[Ball(r_c, s)]$ 为超球体 $Ball(\hat{r}_c, s)$ 体积, $Vol[Cube(2\vartheta, s)]$ 为立方体 $Cube(2\vartheta, s)$ 的体积, ϑ 为特征尺度;

确定等体积超球体法筛选条件, 表达式为:

$$r_c \sum_{i=1}^s \theta_{q,i}^2 \leq (r_c)^2, q = 1, 2, \dots, o_{sel}$$

$$r_c = \begin{cases} \left[\frac{(2\vartheta)^{2k} \cdot k!}{\pi^k} \right]^{1/2k}, & \text{when } s = 2k, k = 1, 2, \dots, s/2 \\ \left[\frac{(2\vartheta)^{2k+1} \cdot \prod_{l=0}^k (2l+1)}{2(2\pi)^k} \right]^{1/(2k+1)}, & \text{when } s = 2k+1, k = 1, 2, \dots, (s-1)/2 \end{cases}$$

其中 $\sum_{i=1}^s \theta_{q,i}^2$ 为第 q 个点在 s 维空间中各维度坐标的平方和, o_{sel} 为被选中的代表点数量。

6. 基于大数据人工智能分析的智能车数据采集系统, 用以执行权利要求1-5任一项所述的方法, 其特征在于, 包括:

数据采集模块:包括GPS传感器、IMU传感器、智能驾驶摄像头、激光雷达、超声波雷达;用于采集行车过程的传感数据和状态数据,对所述传感数据和所述状态数据进行预处理;

图像提取模块:用于对所述环境图像进行图像提取获得驾驶环境数据;

数据处理模块:用于整合所述扫描数据确定行车路况和行车车况,用于筛选所述区域驾驶环境数据异常数据并确定对应时刻状态数据,根据所述状态数据确定车辆状态因子;

数据交互模块:用于将所述用户驾驶环境数据上传至用户驾驶环境数据共享平台,读取并匹配区域驾驶环境数据;

评分模型模块:根据所述区域驾驶环境数据和所述车辆状态因子构建智能车路评分模型,将行车过程的传感数据和状态数据输入所述智能车路评分模型获得车路评分;

智能监管模块:用于储存、查看和管理所述行车过程的传感数据、状态数据和所述车路评分,根据所述车路评分选定智能车数据采集策略并进行数据采集。

基于大数据人工智能分析的智能车数据采集方法及系统

技术领域

[0001] 本发明涉及数据采集领域,尤其涉及基于大数据人工智能分析的智能车数据采集方法及系统。

背景技术

[0002] 随着汽车工业的飞速发展和智能交通系统的兴起,智能车在现代交通领域扮演着越来越重要的角色,而大部分智能车都配备了各种先进的传感器,这些传感器在行车过程中持续产生海量的数据,有效地采集、处理和利用这些数据,对提高行车安全、优化行车效率、减少能源消耗有重要意义。

[0003] 传统的数据采集方法大多依赖于单一的传感器或设备,难以全面捕捉行车过程中的复杂信息,还存在数据处理效率低下、数据价值挖掘不充分、缺乏对不同驾驶环境和车辆状态的综合分析等缺陷。同时,海量数据的存储和管理也面临着巨大挑战,缺乏有效的共享和整合机制,导致不同区域和车辆之间的数据无法充分交互与利用,难以实现智能交通系统的协同优化。因此基于大数据和机器学习技术,设计基于大数据人工智能分析的智能车数据采集方法及系统,克服现有智能车数据采集方法的不足,为智能车的优化控制、智能决策以及智能交通系统的整体规划提供有力的数据依据,对推动智能交通领域的进一步发展有重要意义。

发明内容

[0004] 本发明的目的是要提供基于大数据人工智能分析的智能车数据采集方法及系统。

[0005] 为达到上述目的,本发明是按照以下技术方案实施的:

[0006] 本发明包括以下步骤:

[0007] 获取行车过程的传感数据和状态数据,对所述传感数据和所述状态数据进行预处理;所述传感数据包括环境图像和扫描数据;

[0008] 整合所述扫描数据确定行车路况和行车车况,对所述环境图像进行图像提取获得驾驶环境数据,将所述驾驶环境数据、所述行车路况和所述状态数据组成用户驾驶环境数据;

[0009] 构建用户驾驶环境数据共享平台,上传所述用户驾驶环境数据并匹配区域驾驶环境数据,筛选所述区域驾驶环境数据,根据所述区域驾驶环境数据筛选结果确定车辆状态因子;

[0010] 根据所述区域驾驶环境数据和所述车辆状态因子构建智能车路评分模型,将行车过程的传感数据和状态数据输入所述智能车路评分模型获得车路评分,根据所述车路评分进行智能车数据采集。

[0011] 进一步的,获取行车过程的传感数据和状态数据的方法,包括:

[0012] 获得传感数据的方法包括:通过车身四周摄像头采集行车过程的环境图像,通过车身各处的激光雷达、毫米波雷达和超声波雷达采集行车过程的距离扫描数据;

[0013] 获得状态数据的方法包括：通过GPS传感器采集行车定位数据，通过IMU传感器采集行车运动状态和姿态数据，直接通过车辆系统获取车辆状态数据。

[0014] 进一步的，整合所述扫描数据确定行车路况和行车车况的方法，包括：

[0015] 将扫描数据划分为训练集和测试集；

[0016] 构建行车数据模型，所述行车数据模型包括决策树、特征融合层和堆叠分类器；

[0017] 决策树对输入扫描数据进行特征选择，并将扫描数据划分成行车路况和行车车况两类；

[0018] 特征融合层包括多头注意力机制和特征融合策略，采用多头自注意力机制捕捉不同角度扫描数据间的空间关系、并行处理多个不同类别距离数据，采用Concatenate策略拼接融合不同维度的特征向量，表达式为：

$$[0019] \quad \text{Attention}(Q, K, V) = \frac{\Phi(Q) \sum_{j=1}^n \Phi(K_j)^T V_j}{\Phi(Q) \sum_{j=1}^n \Phi(K_j)^T \sqrt{d_k}} + \lambda \cdot R(\Phi(Q), \Phi(K), \Phi(V))$$

$$[0020] \quad \text{MultiHead}(Q, K, V) = \text{Concat}(\text{head}_1, \text{head}_2, \dots, \text{head}_h) W^0$$

$$[0021] \quad \text{head}_i = \text{Attention}(QW_i^Q, KW_i^K, VW_i^V)$$

[0022] 其中 $\text{Attention}(Q, K, V)$ 为注意力机制中查询向量 Q 在键值对上的注意力输出， K 为关注内容键向量， V 为关注内容值向量， $\Phi(\cdot) = \text{elu}(\cdot) + 1$ 为激活函数， K_j 为第 j 个关注内容键向量， V_j 为第 j 个关注内容值向量， n 为键值对数量， d_k 为尺度化因子， λ 为正则化强度超参数， $R(\cdot)$ 为L2正则化项， $\text{MultiHead}(Q, K, V)$ 为多头注意力机制表示，用于连接多个注意力， W^0 为初始参数矩阵， W_i^Q 为第 i 头查询向量参数矩阵， W_i^K 为第 i 头关注内容键向量参数矩阵， W_i^V 为第 i 头关注内容值向量参数矩阵， head_i 为第 i 头输出；

[0023] 堆叠分类器处理特征融合层输出数据，包括基模型和元模型，基模型间独立学习距离数据进行距离预测，元模型对基模型的距离预测结果进行最终决策，输出行车路况和行车车况；所述基模型包括随机森林、K近邻和支持向量机；

[0024] 采用交叉熵损失函数区分行车路况和行车车况数据，采用网格搜索对行车数据模型的超参数进行调优，采用测试集数据对模型进行评估；

[0025] 将待整合扫描数据输入行车数据模型获得行车路况和行车车况。

[0026] 进一步的，匹配所述区域驾驶环境数据的方法，包括：

[0027] 将驾驶环境数据、行车路况和GPS定位信息组成用户驾驶环境数据；

[0028] 构建用户驾驶环境数据共享平台，用户驾驶环境数据共享平台通过用户数据共享合约进行区域用户间驾驶环境数据的实时交互，采用对称加密算法对用户数据进行加密后传输，采用非对称算法验证用户身份和授权用户读取数据；

[0029] 根据用户驾驶环境数据匹配用户驾驶环境数据共享平台数据库，定义本用户驾驶环境数据向量为 $A = (a_1, a_2, \dots, a_d)$ ，定义数据库中区域其他用户驾驶环境数据向量为 $B_j = (b_{j1}, b_{j2}, \dots, b_{jd})$ ，其中 d 为向量维度，根据数据相似度确定区域驾驶环境数据，表达式为：

$$[0030] \quad Sim_j = \frac{\sum_{i=1}^d (a_i \cdot b_{ji})}{\sqrt{\sum_{i=1}^d (a_i)^2} \sqrt{\sum_{i=1}^d (b_{ji})^2}} \left[\alpha \left(1 - \frac{cov(A, B_j)}{\sigma(A)\sigma(B_j)} \right) + \frac{\beta}{d} \sqrt{\sum_{i=1}^d (a_i - b_{ji})^2} \right]$$

[0031] 其中 Sim_j 为本用户驾驶环境数据向量 A 与区域其他用户驾驶环境数据向量 B_j 的综合相似度, α 为差异波动占比, β 为幅值差异占比, 且 $\alpha + \beta = 1$, $cov(A, B_j)$ 为向量 A 与向量 B_j 的协方差, $\sigma(A)$ 为向量 A 的标准差, $\sigma(B_j)$ 为向量 B_j 的标准差;

[0032] 选取与本用户驾驶环境数据向量 A 综合相似度最高的 m 个区域其他用户驾驶环境数据向量 B_j , 计算区域驾驶环境数据向量 $C = [A + \sum_{j=1}^m B_j] / (1 + m)$, 根据区域驾驶环境数据向量 C 确定区域驾驶环境数据。

[0033] 进一步的, 筛选所述区域驾驶环境数据的方法, 包括:

[0034] 采用等体积超球体法筛选区域驾驶环境数据, 设超球体为 $Ball(r_c, s)$, r_c 为超球体半径, s 为超球体维度, $Vol[Ball(r_c, s)]$ 为超球体体积, 通过搜寻最小星偏差优化等效半径 r_c , 表达式为:

$$[0035] \quad \hat{R}_{EVHS} = arg \left\{ \min_{\hat{r}_c \in r_c} \sup \left| \frac{N[Ball(\hat{r}_c, s), \hat{R}]}{o} - \frac{Vol[Ball(r_c, s)]}{Vol[Cube(2\vartheta, s)]} \right| \right\}$$

[0036] 其中 \hat{R}_{EVHS} 为搜寻最小星偏差得到的均匀性最好的有效代表点集, \hat{R} 为参考点集, \hat{r}_c 为最优等效半径, $N[Ball(\hat{r}_c, s), \hat{R}]$ 为半径为 \hat{r}_c 的超球体 $Ball(\hat{r}_c, s)$ 内有效代表点数量, o 为总的数据点数量, $Vol[Ball(r_c, s)]$ 为超球体 $Ball(\hat{r}_c, s)$ 体积, $Vol[Cube(2\vartheta, s)]$ 为立方体 $Cube(2\vartheta, s)$ 的体积, ϑ 为特征尺度;

[0037] 确定等体积超球体法筛选条件, 表达式为:

$$[0038] \quad r_c \sum_{i=1}^s \theta_{q,i}^2 \leq (r_c)^2, q = 1, 2, \dots, o_{sel}$$

$$[0039] \quad r_c = \begin{cases} \left[\frac{(2\vartheta)^{2k} \cdot k!}{\pi^k} \right]^{1/2k}, & \text{when } s = 2k, k = 1, 2, \dots, s/2 \\ \left[\frac{(2\vartheta)^{2k+1} \cdot \prod_{l=0}^k (2l+1)}{2(2\pi)^k} \right]^{1/(2k+1)}, & \text{when } s = 2k+1, k = 1, 2, \dots, (s-1)/2 \end{cases}$$

[0040] 其中 $\sum_{i=1}^s \theta_{q,i}^2$ 为第 q 个点在 s 维空间中各维度坐标的平方和, o_{sel} 为被选中的代表点数量。

[0041] 进一步的, 根据所述区域驾驶环境数据筛选结果确定车辆状态因子的方法, 包括:

[0042] 根据区域驾驶环境数据筛选结果确定对应时刻的状态数据, 将状态数据输入状态影响函数中获得车辆状态因子, 表达式为:

$$\begin{aligned}
 [0043] \quad \text{Affect} = & w_1 \left[(1 - \sin \sqrt{U/M})(1 - \sin \sqrt{\delta})e^{-D} \right] \left\{ w_2 \left[v_t \ln \left(1 + \frac{v_t - v_{t0}}{v_{t0}} \right) + e^{-g_t} + e^{-\sin P_{pose}} \right] \right. \\
 & \left. + w_3 \sin \left(\sum_{i=1}^{n_1} c_i e^{P_i} + \sum_{j=1}^{n_2} c_j e^{H_j} \right) + \frac{w_4}{T_w} \min (P_{pre}, P_{op}, P_{oq}) \right\}
 \end{aligned}$$

[0044] 其中 *Affect* 为车辆状态因子, w_1 为失控概率权重, w_2 为速度状态因素权重, w_3 为安全驾驶因素权重, w_4 用户使用度权重, U 为辅助驾驶系统用户量, M 为额定用户量, δ 为辅助驾驶系统BUG概率, D 为网络延迟率, v_t 为当前车辆驾驶速度, v_{t0} 为当前时刻街道规定驾驶速度, g_t 为当前车辆加速度, P_{pose} 为当前车辆姿势评分, c_i 为车辆安全保护结构类别, n_1 为车辆安全保护结构类别数量, P_i 为车辆安全保护结构安全评分, c_j 为车辆驾驶影响因素类别, n_2 为车辆驾驶影响因素类别数量, H_j 为当前车辆驾驶影响因素值, H_{j0} 为当前车辆驾驶影响因素标准值, T_w 为当前车辆行驶时间, P_{pre} 为轮胎气压评分, P_{op} 为邮箱油压评分, P_{oq} 为油箱油量评分。

[0045] 进一步的, 根据所述车路评分进行智能车数据采集的方法, 包括:

[0046] 根据所述区域驾驶环境数据和所述车辆状态因子构建智能车路评分模型, 将行车过程的传感数据和状态数据输入所述智能车路评分模型获得车路评分, 根据所述车路评分进行智能车数据采集

[0047] 将区域驾驶环境数据异常数据和对应车辆状态因子组成车路数据集, 将车路数据集划分成训练集和测试集;

[0048] 构建基于孤立森林的智能车路评分模型, 输入层对输入数据进行预处理和特征选择, 孤立森林模型为每个数据点输出异常评分, 输出层根据异常评分和预设阈值为整个数据组分配一个整体评分、提取异常数据标签;

[0049] 采用训练集训练孤立森林模型, 采用交叉熵损失函数进行多类别异常划分, 采用RMSprop优化器调整模型学习率, 采用早停策略控制训练进程, 采用测试集评估模型准确性;

[0050] 将行车过程的传感数据和状态数据输入所述智能车路评分模型获得车路评分和对应的异常标签, 根据车路评分选定智能车数据采集策略。

[0051] 第二方面, 基于大数据人工智能分析的智能车数据采集系统, 包括:

[0052] 数据采集模块: 包括GPS传感器、IMU传感器、智能驾驶摄像头、激光雷达、超声波雷达; 用于采集行车过程的传感数据和状态数据, 对所述传感数据和所述状态数据进行预处理;

[0053] 图像提取模块: 用于对所述环境图像进行图像提取获得驾驶环境数据;

[0054] 数据处理模块: 用于整合所述扫描数据确定行车路况和行车车况, 用于筛选所述区域驾驶环境数据异常数据并确定对应时刻状态数据, 根据所述状态数据确定车辆状态因子;

[0055] 评分模型模块: 根据所述区域驾驶环境数据和所述车辆状态因子构建智能车路评分模型, 将行车过程的传感数据和状态数据输入所述智能车路评分模型获得车路评分;

[0056] 智能监管模块: 用于储存、查看和管理所述行车过程的传感数据、状态数据和所述车路评分, 根据所述车路评分选定智能车数据采集策略并进行数据采集。

[0057] 本发明的有益效果是：

[0058] 本发明是基于大数据人工智能分析的智能车数据采集方法及系统，与现有技术相比，本发明具有以下技术效果：

[0059] 本发明通过数据融合、图像提取、数据交互、异常数据筛选、构建车辆状态因子、构建评分模型步骤，可以在智能车数据采集中提升数据预处理的能力和增强模型适应性，从而提高智能车数据采集的效率，将智能车数据采集技术优化，可以大大节省资源，提高工作效率，实现对智能车的数据采集，对行车效率优化和行车安全提升具有重要意义，可以适应不同基于大数据人工智能分析的智能车数据采集系统、不同用户的基于大数据人工智能分析的智能车的数据采集需求，具有一定的普适性。

附图说明

[0060] 图1为本发明基于大数据人工智能分析的智能车数据采集方法的步骤流程图。

具体实施方式

[0061] 下面通过具体实施例对本发明作进一步描述，在此发明的示意性实施例以及说明用来解释本发明，但并不作为对本发明的限定。

[0062] 本发明基于大数据人工智能分析的智能车数据采集方法及系统包括以下步骤：

[0063] 如图1所示，在本实施例中，包括以下步骤：

[0064] 获取行车过程的传感数据和状态数据，对所述传感数据和所述状态数据进行预处理；所述传感数据包括环境图像和扫描数据；

[0065] 整合所述扫描数据确定行车路况和行车车况，对所述环境图像进行图像提取获得驾驶环境数据，将所述驾驶环境数据、所述行车路况和所述状态数据组成用户驾驶环境数据；

[0066] 构建用户驾驶环境数据共享平台，上传所述用户驾驶环境数据并匹配区域驾驶环境数据，筛选所述区域驾驶环境数据，根据所述区域驾驶环境数据筛选结果确定车辆状态因子；

[0067] 根据所述区域驾驶环境数据和所述车辆状态因子构建智能车路评分模型，将行车过程的传感数据和状态数据输入所述智能车路评分模型获得车路评分，根据所述车路评分进行智能车数据采集。

[0068] 在本实施例中，获取行车过程的传感数据和状态数据的方法，包括：

[0069] 获得传感数据的方法包括：通过车身四周摄像头采集行车过程的环境图像，通过车身各处的激光雷达、毫米波雷达和超声波雷达采集行车过程的距离扫描数据；

[0070] 获得状态数据的方法包括：通过GPS传感器采集行车定位数据，通过IMU传感器采集行车运动状态和姿态数据，直接通过车辆系统获取车辆状态数据；所述车辆状态数据包括安全带及安全气囊状态、轮胎胎压、油箱油量、行驶挡位、来车向灯光和驾驶噪声。

[0071] 在本实施例中，整合所述扫描数据确定行车路况和行车车况的方法，包括：

[0072] 将扫描数据划分为训练集和测试集；

[0073] 构建行车数据模型，所述行车数据模型包括决策树、特征融合层和堆叠分类器；

[0074] 决策树对输入扫描数据进行特征选择，并将扫描数据划分成行车路况和行车车况

两类;所述行车路况包括车辆与四周道路、路口的距离;所述行车车况包含车辆与障碍物、四周车辆、护栏的距离;

[0075] 特征融合层包括多头注意力机制和特征融合策略,采用多头自注意力机制捕捉不同角度扫描数据间的空间关系、并行处理多个不同类别距离数据,采用Concatenate策略拼接融合不同维度的特征向量,表达式为:

$$[0076] \quad \text{Attention}(Q, K, V) = \frac{\Phi(Q) \sum_{j=1}^n \Phi(K_j)^T V_j}{\Phi(Q) \sum_{j=1}^n \Phi(K_j)^T \sqrt{d_k}} + \lambda \cdot R(\Phi(Q), \Phi(K), \Phi(V))$$

$$[0077] \quad \text{MultiHead}(Q, K, V) = \text{Concat}(\text{head}_1, \text{head}_2, \dots, \text{head}_h) W^0$$

$$[0078] \quad \text{head}_i = \text{Attention}(QW_i^Q, KW_i^K, VW_i^V)$$

[0079] 其中 $\text{Attention}(Q, K, V)$ 为注意力机制中查询向量 Q 在键值对上的注意力输出, K 为关注内容键向量, V 为关注内容值向量, $\Phi(\cdot) = \text{elu}(\cdot) + 1$ 为激活函数, K_j 为第 j 个关注内容键向量, V_j 为第 j 个关注内容值向量, n 为键值对数量, d_k 为尺度化因子, λ 为正则化强度超参数, $R(\cdot)$ 为L2正则化项, $\text{MultiHead}(Q, K, V)$ 为多头注意力机制表示,用于连接多个注意力, W^0 为初始参数矩阵, W_i^Q 为第 i 头查询向量参数矩阵, W_i^K 为第 i 头关注内容键向量参数矩阵, W_i^V 为第 i 头关注内容值向量参数矩阵, head_i 为第 i 头输出;

[0080] 堆叠分类器处理特征融合层输出数据,包括基模型和元模型,基模型间独立学习距离数据进行距离预测,元模型对基模型的距离预测结果进行最终决策,输出行车路况和行车车况;所述基模型包括随机森林、K近邻和支持向量机;

[0081] 采用交叉熵损失函数区分行车路况和行车车况数据,采用网格搜索对行车数据模型的超参数进行调优,采用测试集数据对模型进行评估;

[0082] 将待整合扫描数据输入行车数据模型获得行车路况和行车车况;

[0083] 在实际评估中,获取某汽车用户行车过程的传感数据,将行车过程的传感数据输入训练好的行车数据模型获得:智能车ID:CAR12***88,时间戳:2024-06-13T17:22:01Z,行车路况(车辆与前方路口的距离150.5m、与道路边缘的距离2.1m、与交通信号灯的距离15m、与中间护栏距离3.8m),行车车况(左侧车辆相距0.8m-速度30km/h、前方车辆相距15.2m-速度40km/h、后方车辆相距20.1m-速度42km/h、前方障碍物相距5.6m、前方斑马线行人相距10.3m)。

[0084] 在本实施例中,对所述环境图像进行图像提取获得驾驶环境数据的方法,包括:

[0085] 采用随机森林算法将环境图像按7:3划分成训练集和测试集;

[0086] 构建驾驶环境图像模型,驾驶环境图像模型包括CNN卷积神经网络、全连接层、目标检测算法、图像分割技术、堆叠泛化和决策树;

[0087] CNN卷积神经网络对环境图像进行深度学习提取图像特征,将特征提取结果输入后置结构;

[0088] 后置结构为并联排列,包括全连接层、目标检测算法、图像分割技术,全连接层使用softmax激活函数识别环境图像中的红绿灯、交通标识、路标和危险源等特征,使用Focal Loss损失函数衡量减少易分类样本的损失贡献,增加难分类样本的损失贡献;采用Faster

R-CNN目标检测算法提取信号灯时间、交通标识数据、人流量、车流量和危险源数量,采用平均绝对误差调整预测值和真实值的差异;图像分割技术区分危险场景和天气的场景元素,采用多类别交叉熵损失函数调整多标签分类场景;

[0089] 堆叠泛化将特征识别、数据提取和场景识别的结果进行拼接融合获得综合识别结果,采用决策树将驾驶环境数据识别结果进行类别划分,采用结构损失函数优化决策树结构;所述道路信号包括红绿灯状态、交通标识和路标;所述道路流量包括人流量和车流量;所述道路危险包括危险源类型、危险源数量和危险场景;

[0090] 采用网格搜索技术对模型超参数进行调优,采用测试集评估模型的准确性;

[0091] 将待处理的环境图像输入驾驶环境图像模型获得驾驶环境数据;

[0092] 在实际评估中,采用驾驶环境图像模型处理该车辆环境图像得到:道路信号(前方绿灯、信号灯剩余12s、路标限速45km/h、前方学校、双车道),道路流量(前方斑马线通行8人/行人密集、右前方非机动车道10辆非机动车、右后方8辆非机动车、前方视线12辆车/轻微拥挤、后方视线6辆车/畅通),道路危险(障碍物-临时故障标志、障碍物数量1个、危险源-前方20m有满载大型货车、无危险事故或场景),实况天气(小雨、地面湿滑、能见度良好)。

[0093] 在本实施例中,匹配所述区域驾驶环境数据的方法,包括:

[0094] 将驾驶环境数据、行车路况和GPS定位信息组成用户驾驶环境数据;

[0095] 构建用户驾驶环境数据共享平台,用户驾驶环境数据共享平台通过用户数据共享合约进行区域用户间驾驶环境数据的实时交互,采用对称加密算法对用户数据进行加密后传输,采用非对称算法验证用户身份和授权用户读取数据;

[0096] 根据用户驾驶环境数据匹配用户驾驶环境数据共享平台数据库,定义本用户驾驶环境数据向量为 $A = (a_1, a_2, \dots, a_d)$, 定义数据库中区域其他用户驾驶环境数据向量为 $B_j = (b_{j1}, b_{j2}, \dots, b_{jd})$, 其中 d 为向量维度, 根据数据相似度确定区域驾驶环境数据, 表达式为:

$$[0097] \quad Sim_j = \frac{\sum_{i=1}^d (a_i \cdot b_{ji})}{\sqrt{\sum_{i=1}^d (a_i)^2} \sqrt{\sum_{i=1}^d (b_{ji})^2}} \left[\alpha \left(1 - \frac{cov(A, B_j)}{\sigma(A)\sigma(B_j)} \right) + \frac{\beta}{d} \sqrt{\sum_{i=1}^d (a_i - b_{ji})^2} \right]$$

[0098] 其中 Sim_j 为本用户驾驶环境数据向量 A 与区域其他用户驾驶环境数据向量 B_j 的综合相似度, α 为差异波动占比, β 为幅值差异占比, 且 $\alpha + \beta = 1$, $cov(A, B_j)$ 为向量 A 与向量 B_j 的协方差, $\sigma(A)$ 为向量 A 的标准差, $\sigma(B_j)$ 为向量 B_j 的标准差;

[0099] 选取与本用户驾驶环境数据向量 A 综合相似度最高的 m 个区域其他用户驾驶环境数据向量 B_j , 计算区域驾驶环境数据向量 $C = [A + \sum_{j=1}^m B_j] / (1 + m)$, 根据区域驾驶环境数据向量 C 确定区域驾驶环境数据;

[0100] 在实际评估中,将该用户驾驶环境数据上传用户驾驶环境数据共享平台并获取附近区域用户驾驶环境数据:位于用户前方18m/左车道,行车路况(车辆与前方路口的距离-未知、与道路边缘的距离5.6m、与交通信号灯的距离100m、与中间护栏距离0.3m),行车车况(右侧车辆相距0.8m-速度36km/h、前方车辆相距20m-速度12km/h、后方车辆相距10m-速度25km/h、无障碍物),道路信号(前方绿灯、信号灯剩余22s、路标限速20km/h、前方学校、双车

道),道路流量(前方斑马线通行25人/行人拥挤、右前方非机动车道22辆非机动车、右后方19辆非机动车、前方视线18辆车/拥堵、后方视线10辆车/轻微拥堵),道路危险(无障碍物、危险源右侧有满载大型货车、无危险事故或场景),实况天气(小雨、地面湿滑、能见度良好)。

[0101] 在本实施例中,根据筛选所述区域驾驶环境数据的方法,包括:

[0102] 采用等体积超球体法筛选区域驾驶环境数据,设超球体为 $Ball(r_c, s)$, r_c 为超球体半径, s 为超球体维度, $Vol[Ball(r_c, s)]$ 为超球体体积,通过搜寻最小星偏差优化等效半径 r_c , 表达式为:

$$[0103] \quad \hat{R}_{EVHS} = arg \left\{ \min_{\hat{r}_c \in r_c} \sup \left| \frac{N[Ball(\hat{r}_c, s), \hat{R}]}{o} - \frac{Vol[Ball(r_c, s)]}{Vol[Cube(2\vartheta, s)]} \right| \right\}$$

[0104] 其中 \hat{R}_{EVHS} 为搜寻最小星偏差得到的均匀性最好的有效代表点集, \hat{R} 为参考点集, \hat{r}_c 为最优等效半径, $N[Ball(\hat{r}_c, s), \hat{R}]$ 为半径为 \hat{r}_c 的超球体 $Ball(\hat{r}_c, s)$ 内有效代表点数量, o 为总的点数数量, $Vol[Ball(r_c, s)]$ 为超球体 $Ball(\hat{r}_c, s)$ 体积, $Vol[Cube(2\vartheta, s)]$ 为立方体 $Cube(2\vartheta, s)$ 的体积, ϑ 为特征尺度;

[0105] 确定等体积超球体法筛选条件,表达式为:

$$[0106] \quad r_c \sum_{i=1}^s \theta_{q,i}^2 \leq (r_c)^2, q = 1, 2, \dots, o_{sel}$$

$$[0107] \quad r_c = \begin{cases} \left[\frac{(2\vartheta)^{2k} \cdot k!}{\pi^k} \right]^{1/2k}, & \text{when } s = 2k, k = 1, 2, \dots, s/2 \\ \left[\frac{(2\vartheta)^{2k+1} \cdot \prod_{l=0}^k (2l+1)}{2(2\pi)^k} \right]^{1/(2k+1)}, & \text{when } s = 2k+1, k = 1, 2, \dots, (s-1)/2 \end{cases}$$

[0108] 其中 $\sum_{i=1}^s \theta_{q,i}^2$ 为第 q 个点在 s 维空间中各维度坐标的平方和, o_{sel} 为被选中的代表点数量;

[0109] 在实际评估中,该区域驾驶环境数据中的异常数据为:用户车辆前方斑马线行人通行、前方5.6m处有障碍物、前方20m有满载/未减速大型货车、小雨且地面湿滑、前方学校行人道路拥堵。

[0110] 在本实施例中,根据所述区域驾驶环境数据筛选结果确定车辆状态因子的方法,包括:

[0111] 根据区域驾驶环境数据筛选结果确定对应时刻的状态数据,将状态数据输入状态影响函数中获得车辆状态因子,表达式为:

$$[0112] \quad Affect = w_1 \left[(1 - \sin \sqrt{U/M})(1 - \sin \sqrt{\delta})e^{-D} \right] \left\{ w_2 \left[v_t \ln \left(1 + \frac{v_t - v_{t0}}{v_{t0}} \right) + e^{-\vartheta t} + e^{-\sin P_{poss}} \right] \right. \\ \left. + w_3 \sin \left(\sum_{i=1}^{n_1} c_i e^{P_i} + \sum_{j=1}^{n_2} c_j e^{\frac{H_j}{H_{j0}}} \right) + \frac{w_4}{T_w} \min (P_{pre}, P_{opp}, P_{oq}) \right\}$$

[0113] 其中 $Affect$ 为车辆状态因子, w_1 为失控概率权重, w_2 为速度状态因素权重, w_3 为安全驾驶因素权重, w_4 用户使用度权重, U 为辅助驾驶系统用户量, M 为额定用户量, δ 为辅

助驾驶系统BUG概率, D 为网络延迟率, v_t 为当前车辆驾驶速度, v_{t0} 为当前时刻街道规定驾驶速度, g_t 为当前车辆加速度, P_{pose} 为当前车辆姿势评分, c_i 为车辆安全保护结构类别, n_1 为车辆安全保护结构类别数量, P_i 为车辆安全保护结构安全评分, c_j 为车辆驾驶影响因素类别, n_2 为车辆驾驶影响因素类别数量, H_j 为当前车辆驾驶影响因素值, H_{j0} 为当前车辆驾驶影响因素标准值, T_w 为当前车辆行驶时间, P_{pre} 为轮胎气压评分, P_{op} 为邮箱油压评分, P_{oq} 为油箱油量评分;

[0114] 在实际评估中, 车辆状态因子权重取 $w_1 = w_2 = w_3 = w_4 = 0.25$, 将区域驾驶环境数据异常数据对应时刻的状态数据(

$$U = 150k, M = 200k, \delta = 0.02\%, D = 0.05, v_t = 48km/h, v_0 = 45km/h,$$

$h, g_t = -0.5m/m^2, P_{pose} = 0.8, T_w = 1h, n_1 = 3, n_2 = 2, c_i = (1,1,1), P_i = (2,3,4), c_j = (1,1), H_j = (0.05,0.1), H_{j0} = (0.05,0.1)$) 输入状态影响函数

中获得车辆状态因子 $Affect = 1.135$ 。

[0115] 在本实施例中, 根据所述车路评分进行智能车数据采集的方法, 包括:

[0116] 根据所述区域驾驶环境数据和所述车辆状态因子构建智能车路评分模型, 将行车过程的传感数据和状态数据输入所述智能车路评分模型获得车路评分, 根据所述车路评分进行智能车数据采集

[0117] 将区域驾驶环境数据异常数据和对应车辆状态因子组成车路数据集, 将车路数据集划分成训练集和测试集;

[0118] 构建基于孤立森林的智能车路评分模型, 输入层对输入数据进行预处理和特征选择, 孤立森林模型为每个数据点输出异常评分, 输出层根据异常评分和预设阈值为整个数据组分配一个整体评分、提取异常数据标签;

[0119] 采用训练集训练孤立森林模型, 采用交叉熵损失函数进行多类别异常划分, 采用RMSprop优化器调整模型学习率, 采用早停策略控制训练进程, 采用测试集评估模型准确性;

[0120] 将行车过程的传感数据和状态数据输入所述智能车路评分模型获得车路评分和对应的异常标签, 根据车路评分选定智能车数据采集策略; 当车路评分大于设定阈值时采集对应时间点行车过程状态数据, 当车路评分小于设定阈值时采集对应时间点行车过程传感数据和状态数据;

[0121] 在实际评估中, 将行车过程的传感数据和状态数据输入所述智能车路评分模型获得车路评分为6.6分、异常情况(包括斑马线行人通行、障碍物、满载未减速同行大型货车、小雨且地面湿滑、学校路段、行人道路拥堵), 车路评分小于车路评分阈值8分, 采集该时段行车过程传感数据和状态数据。

[0122] 第二方面, 基于大数据人工智能分析的智能车数据采集系统, 包括:

[0123] 数据采集模块: 包括GPS传感器、IMU传感器、智能驾驶摄像头、激光雷达、超声波雷达; 用于采集行车过程的传感数据和状态数据, 对所述传感数据和所述状态数据进行预处理;

[0124] 图像提取模块: 用于对所述环境图像进行图像提取获得驾驶环境数据;

[0125] 数据处理模块：用于整合所述扫描数据确定行车路况和行车车况，用于筛选所述区域驾驶环境数据异常数据并确定对应时刻状态数据，根据所述状态数据确定车辆状态因子；

[0126] 评分模型模块：根据所述区域驾驶环境数据和所述车辆状态因子构建智能车路评分模型，将行车过程的传感数据和状态数据输入所述智能车路评分模型获得车路评分；

[0127] 智能监管模块：用于储存、查看和管理所述行车过程的传感数据、状态数据和所述车路评分，根据所述车路评分选定智能车数据采集策略并进行数据采集。

[0128] 以上所述仅为本发明的较佳实施例，并不用以限制本发明，凡在本发明的精神和原则之内，所作的任何修改、等同替换、改进等，均应包含在本发明的保护范围之内。

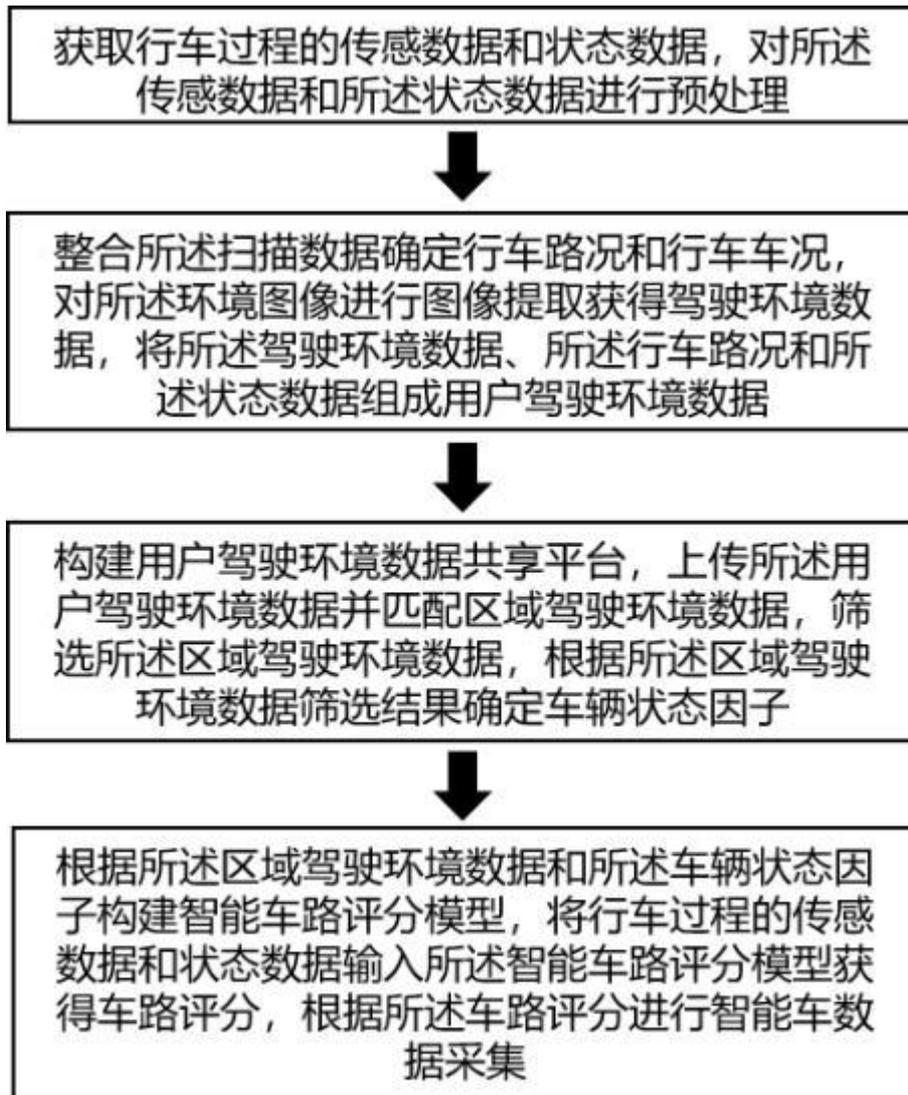


图 1