

“新国线杯”

第十五届全国大学生交通运输科技大赛

“交通大数据” 竞赛类竞赛方案

一、竞赛意义

“交通大数据”主题竞赛利用实际城市交通大数据，搭建交通信息化发展和应用探索与实践的平台，开展多方面交通数据挖掘研究。目的在于：

（1）推动基于大数据的新一代智能交通系统发展，从数据出发，立足现实，结合理论研究，推进大数据各项工作；

（2）建立一个开放的公共服务支撑平台，提供共享数据、分析工具软件、评价标准及技术、热点问题讨论等；

（3）用大数据提升行业治理能力，用数据说话、用数据管理、用数据决策、用数据创新，围绕现阶段交通行业痛点难点，寻找解决交通实际问题的新思路、新方法。

二、竞赛规则

“交通大数据”主题竞赛要求参赛项目组登录平台，依据大赛主题指定的研究方向和研究内容，从平台获取特定时段的全量轨迹数据，通过测量平台提供的交通信息采集方式，根据其研究需求部署研究环境、获取相关数据、开展相关研究，在规定时间内完成研究并提交作品。

1、数据来源

大赛数据为 OpenITS 联盟提供的安徽宣城特定时段的全量轨迹脱敏数据，时间维度为 30 天，数据采集区域覆盖宣城全部干道，主要包括：

- 任意位点、任意时间、任意统计频率的车道级线圈数据；
- 给定最大渗透率下任意占比的浮动车 GPS 数据。

数据说明：

虚拟测量平台可为大赛提供 30 天的全域、全量、全时数据，主

要提供两类交通信息采集方式：定点检测线圈采集和 GPS 浮动车移动采集。

线圈数据：根据用户在虚拟交通测量平台设置的时间范围、空间位置、以及线圈装置的漏检率和统计间隔，可以输出虚拟线圈所在断面的流量数据和平均车速数据。其数据结构如下：

| 字段名 | 描述 |
|--------------------------|-----------|
| DEVICEID | 设备 ID |
| FROMTIME | 统计起始时间 |
| TOTIME | 统计结束时间 |
| INTERVAL | 统计间隔 |
| LANEID | 车道 ID |
| COUNT | 统计间隔内车辆计数 |
| FLOW | 断面换算小时流量 |
| ARITHMETIC_AVERAGE_SPEED | 车速算术平均值 |
| HARMONIC_AVERAGE_SPEED | 车速调和平均值 |

浮动车 GPS 数据：根据用户在虚拟交通测量平台设定的时空范围内所设置的浮动车渗透率及其 GPS 采样间隔，获取浮动车坐标数据。其数据结构如下：

| 字段名 | 描述 |
|-----------|-------------------|
| TRAVELID | 车辆 ID |
| sj | 检测时间 |
| longitude | 经度 |
| latitude | 纬度 |
| angle | 车辆方向角 |
| speed | 瞬时车速 |
| posm | 该有向路段中相对终止节点的里程位置 |
| ftnode | 浮动车所处路段的起始节点和终止节点 |

地图数据：用户可在数据平台下载对应的地图数据，其数据结构如下：

| 字段名 | 描述 |
|---------|----------|
| ROADID | 路段 ID |
| FNODE | 路段起始节点编号 |
| TNODE | 路段结束节点编号 |
| Geom | 路段地理信息字段 |
| LaneDir | 车道转向信息 |

数据来源网址 <http://nactrans.openits.cn>，数据详情见附件。

2、竞赛选题

“交通大数据”主题竞赛选题分为两个部分，可任选其一完成：

(1) 断面车道级流量预测，预测间隔为 5 分钟

参赛者通过测量平台统一提供的线圈、浮动车测量等方式获取数据，建立并优化交通流预测模型。完成模型训练后，上传模型代码，评分系统将针对给定的 5 个路段断面，分别使用测试数据集对模型进行检测，综合考虑模型 5 个预测结果的预测精度等指标，给出综合评分。

(2) 路段车流量预测，预测间隔为 15 分钟

参赛者通过测量平台统一提供的线圈、浮动车测量等方式获取数据，建立路段车流量预测模型，对不同道路路段进行预测。上传模型代码后，评分系统将针对给定的 5 个路段，并使用测试数据集对模型进行评价，综合考虑模型的预测误差指标，给出综合评分。

3、提交成果形式

为了便于评委了解作品及其技术细节，在作品提交截止日期前，各团队需要提供以下内容：

- (1) 模型完整代码，包括模型的最终参数；
- (2) 结题报告；

结题报告须对以下方面内容进行详细说明：

1. 研究背景：介绍研究的由来、意义、环境、状态、前人研究成果以及当前具有的条件等基础信息。
2. 研究数据支撑：介绍作品所使用到的比赛数据，并详细说明这些数据的使用情况。
3. 研究内容：对作品研究的内容进行梳理介绍。
4. 研究技术路线：详细介绍研究的方法、工具和手段，说明研究的流程
5. 研究结果：对研究的结果进行完整、详细、清晰表达。
6. 研究总结：对研究的问题、过程及成果进行总结。

4、评比规则

大赛将提供 30 天的两个平峰时段（11:00 – 13:00 与 20:00 – 22:00）、两个高峰时段（7:00 – 9:00 与 17:00 – 19:00）的数据作为 **训练集**；并提供 1 周 7 天同样时段的数据作为 **测试集**。参赛者须基

于此完成作品。（训练集见附件，拟指定 5 个断面和 5 条路段进行研究）

为体现大赛的权威性、公正性、公平性，参赛作品的评审方法为：

参赛作品将以算法复杂度和预测结果精度作为主要指标进行量化考核（分数占比 50%），预测结果精度评价指标主要包括如下几个。

$$RMSE = \sqrt{\frac{1}{n} \sum_{i=T+1}^{T+n} (\hat{Y}_i - Y_i)^2}$$

- 均方根误差（RMSE），可对预测效果进行综合评价。
- 绝对误差平均（MAE），可对预测的绝对误差进行综合评价。
- 平均绝对百分误差（MAPE），可对预测的相对误差进行综合评价。

$$MAE = \frac{1}{n} \sum_{i=T+1}^{T+n} |\hat{Y}_i - Y_i|$$

$$MAPE = \frac{1}{n} \sum_{i=T+1}^{T+n} \left| \frac{\hat{Y}_i - Y_i}{Y_i} \right| \times 100$$

除了上述量化评价指标，作品报告完整性（分数占比 10%）、模型算法先进性（分数占比 20%）、研究方法的创新性（分数占比 20%），也将作为参赛作品考核评价的考虑因素。

三、竞赛组织

“交通大数据”主题竞赛的参赛要求、组织程序和评审办法参照《“新国线杯”第十五届全国大学生交通运输科技大赛实施方案》执行。

四、其他说明

1、多场景虚拟交通测量平台相关问题详见附件使用说明，更多问题请咨询：

陈志军老师，联系电话：15827127655，邮箱：
chenzj556@whut.edu.cn

黄玮老师，联系电话：19868388335，邮箱：
huangwei5@mail.sysu.edu.cn

2、未尽事宜，解释权及修改权归大赛组委会。

附件:

OpenITS 多场景虚拟交通测量平台使用说明

随着信息技术的高速发展,交通数据获取能力越来越强、体量越来越大,这些海量的数据为交通领域的研究提供了新的手段和数据基础。交通流信息的检测作为交通系统重要的数据资源,可用于掌控道路交通的运行情况,精确诱导控制等,检测数据的完备性可提升交通数据感知质量,为交通系统的可视、可测、可控提供了技术支撑。

为了契合交通信息技术发展潮流,“OpenITS 研究计划联盟”开放了一个多场景虚拟交通测量平台(<http://vsensor.openits.cn>)。在此测量平台下,可自定义交通测量场景,生成模拟检测数据,以支撑交通基础问题研究。虚拟交通测量平台旨在整合一个全域、全量、全时数据环境,精确刻画车辆出行、路口及路网状态,为路网交通状态感知与分析提供良好基础,从而为本次大赛提供了一个统一的、开放的、可拓展的研究环境。

1、平台简介

多场景虚拟交通测量平台是基于精准重现的全量交通行为仿真平台搭建的交通流参数测量平台,包括线圈和浮动车两类时空范围可选、参数可设置的检测场景。通过开放测量场景,可支撑数据预处理、交通信息提取、数据挖掘、交通状态分析与预测、数据驱动的新型交通理论方法等交通基础问题研究。此外,平台也支撑检测布局优化的基础研究,受限于检测器布设的成本及精度问题,现实中往往无法实现交通网络状态的全面感知。借助虚拟交通测量平台,参赛者在比赛中能在全网范围自主选择各类任意数量的检测手段对网络进行感知,并基于检测数据自定义地布局和开展赛题相关的应用研究。

2、平台说明

虚拟交通测量平台的交通数据主要来源于安徽省宣城市。该城市路网拥有密布的电警卡口,检测粒度可细化至车辆个体级、车道级、秒级的过车信息。虚拟交通测量平台基于交通流冲击波理论,倒推中间运动过程,实现了个体车辆轨迹的重构。该平台正是通过这些完备的出行信息检测与时空轨迹重构,实现了交通场景的全息重构与生成。



图 1 多场景交通测量平台（安徽宣城）

虚拟交通测量平台提供了感应线圈和浮动车 GPS 信息检测等各类测量工具，可以模拟交通信息数据采集，为参赛者提供了自定义检测的研究数据。参赛者可以基于丰富的信息环境，在大赛主题下延伸进行信息与控制、规划与管理、环境与安全等方面的研究。测量平台支撑给定研究课题的比赛，更鼓励参赛者主动发现潜在研究问题、深入分析开放数据集、探索有效解决方法，从而自主开展开放式创新课题研究。

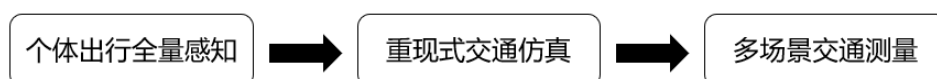


图 2 平台原理与业务

3、数据说明

虚拟测量平台可为大赛提供 30 天的全域、全量、全时数据，主要提供两类交通信息采集方式：定点检测线圈采集和 GPS 浮动车移动采集。参赛团队可根据需要提取不同时段、不同程度或不同检测率的数据集。以下是对平台各类检测数据的说明。

- 线圈数据：根据用户在虚拟交通测量平台设置的时间范围、空间位置、以及线圈装置的漏检率和统计间隔，可以输出虚拟线圈所在断面的流量数据和平均车速数据。其数据结构如下：

| 字段名 | 描述 |
|----------|-----------|
| DEVICEID | 设备 ID |
| FROMTIME | 统计起始时间 |
| TOTIME | 统计结束时间 |
| INTERVAL | 统计间隔 |
| LANEID | 车道 ID |
| COUNT | 统计间隔内车辆计数 |

| | |
|--------------------------|----------|
| FLOW | 断面换算小时流量 |
| ARITHMETIC_AVERAGE_SPEED | 车速算术平均值 |
| HARMONIC_AVERAGE_SPEED | 车速调和平均值 |

- 浮动车 GPS 数据：根据用户在虚拟交通测量平台设定的空间范围内所设置的浮动车渗透率及其 GPS 采样间隔，平台输出模拟 GPS 漂移噪声的浮动车坐标数据。其数据结构如下：

| 字段名 | 描述 |
|-----------|-------------------|
| TRAVELID | 车辆 ID |
| sj | 检测时间 |
| longitude | 经度 |
| latitude | 纬度 |
| angle | 车辆方向角 |
| speed | 瞬时车速 |
| posm | 该有向路段中相对终止节点的里程位置 |
| ftnode | 浮动车所处路段的起始节点和终止节点 |

4、用户操作流程

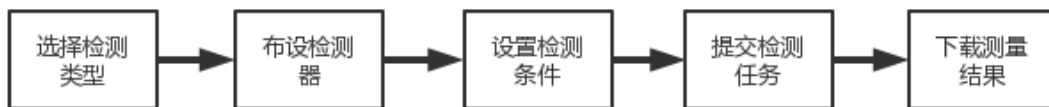


图 3 用户操作流程

- 选择检测类型

用户在添加检测器前，要先在平台页面选择检测器类型。点击“检测器”标签，即可选择检测类型。可供选择的检测类型有两种：线圈和浮动车 GPS 信息检测。

图 4 布置线圈设备



图 5 设置浮动车参数

- 布设检测器

检测器的布设，在本虚拟交通测量平台中主要是指线圈这一断面型检测设备的布设，其包含线圈的增加、修改、删除三方面的内容。

- (1) 增加线圈检测器

用户可通过鼠标交互的方式在车道级地图上选位置打点，以图标来标识生成的断面检测器。用户设置完线圈的设备属性后，点击“保存”按钮则可成功添加设置线圈设备。

- (2) 修改线圈检测器属性

用户在线圈检测器属性窗口更改相关参数后，点击“保存”按钮，即可成功修改检测器属性。

- (3) 删除线圈检测器

用户可通过鼠标交互的方式在车道级地图上点选这些断面型检测设备，再点击设备属性窗口中的“拆除”按钮，即可删除该检测设备。

- 设置检测条件

用户可参考上图 1、2 的窗口界面，来设置不同检测器的检测条件：对于线圈检测器，可设置缺失率和统计间隔；对于浮动车 GPS 信息检测，可设置采样率和渗透率。在设置完检测器的检测条件后，点击各类设备属性窗口中的“保存”按钮，即可成功修改检测器属性。

- 提交检测任务

用户可编辑测量任务的参数。这些操作包括时间范围的选择、空间范围的选择、提交测量任务。

- (1) 时间范围的选择



图 6 选择时间范围

时间范围的选择，即设置测量开始时间、结束时间，最后确认提交，等待测量结果返回。

(2) 空间范围的选择

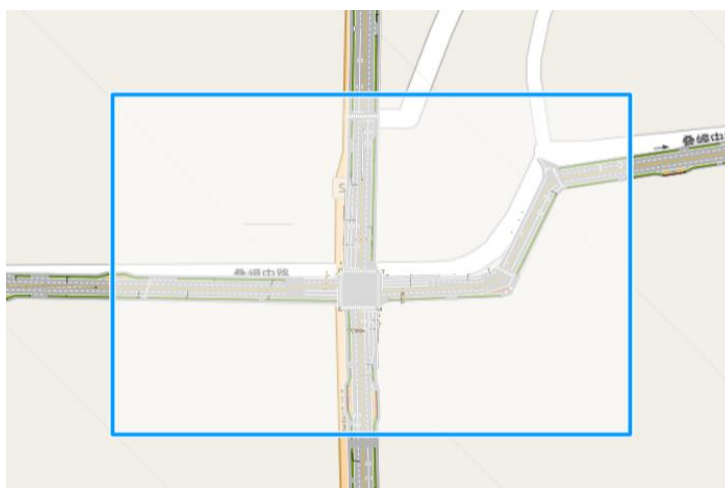


图 7 选择空间范围

用户可利用鼠标拉出矩形选框，测量空间范围的框选，即完成空间选择。其中，用户获得的浮动车 GPS 信息就是来自于该框选的空间范围中。

(3) 提交测量任务



图 8 提交测量任务

用户完成时间范围和空间范围的选择操作后，在“仿真测量”窗口中，点击“提交任务”按钮，即可看到当前任务的状态和执行进度。

- 下载测量结果



图 9 下载测量数据

在“仿真测量”窗口的任务状态栏中，状态显示为“完成任务”后，点击“下载”按钮，即可根据检测请求，对仿真车辆进行测量统计，并返回各类检测结果。

5、平台注册说明

城市交通虚拟测量平台 (vsensor) 目前采用填报方式开放注册。有意参加“交通大数据”竞赛组的高校请填写下表 1，向组委会提出账号开通申请。

原则上每个学校参赛队伍不超过 5 组，若有需要注册更多队伍请与组委会说明情况。

表 1：注册报名表样式

| 队伍编号 | 成员姓名 | 学校名称 | 指导老师 | 联系人姓名及电话 |
|------|-----------|------|------|-----------------------|
| 1 | 成员 1 (队长) | | | 成员 1 (12345678910) |
| | 成员 2 | | | |
| | ... | | | |
| 2 | | | | |
| | | | | |
| | | | | |

6、用户操作限制说明

参赛者对平台的使用将受如下限制:

- 1、线圈检测器布设数量不多于 20 个;
- 2、浮动车渗透率不高于 10%;
- 3、每日下载任务提交次数不多于 10 次,且需要在平台提供的远程计算环境(jupyter)下进行数据处理。