



2023“SEED”第四届江苏大数据开发与应用大赛-智慧交通赛道

参赛队伍

282

进行中

奖金/奖品

220000

我的团队

结果提交

赛事动态

赛事详情

报名指南

赛题详情

一、赛题描述

本赛题将人工智能技术应用于智慧交通领域，以改善交通拥堵问题。赛题聚焦于路口饱和度和快速路OD出行需求的智能化预测，通过分析道路基础数据和交通流数据，预测未来路口交通流饱和情况，辅助实现信控路口交通状态的自动巡检，自动发现存在问题的信控路口，科学分析路口配时方案是否合理，指导信号配时工作。同时，预测快速路OD出行需求，准确掌握大流量OD对，提前分流交通量，从源头上缓解快速路交通拥堵。本赛题鼓励参赛者运用多样化的技术手段，如数据分析、机器学习、深度学习等，提出富有创新性的解决方案，以改善城市交通状况和提升居民出行体验，为城市交通管理和出行规划提供重要参考，促进智慧城市的建设。

二、赛题题目

智慧交通-人工智能在交通领域的应用

三、赛题任务

【初赛】

依托无锡市滨湖区某路口电警卡口的多维脱敏数据，根据路口路况进行特征处理并构造算法模型，预测未来某天每5分钟的路口交通饱和度。

饱和度：实际交通流量（V）与最大通行能力（C）的比值。

最大通行能力C的参考数值：1650。单位：车流量/小时/车道。

结果文件提交时间：2023年10月10日 - 2023年11月8日

【初赛复核】

前50名队伍代码提交时间：2023年11月13日 - 2023年11月25日。初赛代码截止提交时间2023年11月15日。

【复赛】

根据无锡市快速路电警卡口的多维脱敏数据和电子地图数据，预测未来某天一段时间内通过快速路上的不同OD对之间的车辆流量。复赛分为A、B榜。

时间：待定

【决赛】

决赛阶段采用线下答辩的方式，晋级决赛队伍需要提前准备答辩PPT及相关支撑材料，评委将根据选手的初复赛成绩及现场表现进行综合评分，决定最终排名。

时间：暂定

四、数据描述

【初赛】

本赛题提供的数据集包含4张数据表。其中，Lane.csv 为车道数据，Light\_status.csv为路口灯态相位数据，Entrance\_road.csv为进口道数据，Flow.csv为路口车流量数据。数据集为历史27天的数据。

数据集清单与格式说明：

表名	属性	描述
Lane.csv	LANE_ID	车道 id
	ARM_ID	进口道 id
	LANE_NAME	车道转向
	DIR	方向
Light_status.csv	CYCLE_START_TIME	周期开始时间
	STAGE_START_TIME	阶段开始时间
	STAGE_END_TIME	阶段结束时间
	STAGE_LENGTH	阶段总时长
	GREEN_TIME	绿灯时间
	GREEN_FLASH_TIME	绿闪时间
	YELLOW_TIME	黄灯时间
	ALL_RED_TIME	全红时间
	CHANNELS	放行的通道号
	LANES	放行的车道号, 多个以逗号隔开（见注 1）
Entrance_road.csv	ARM_ID	进口道 id
	ARM_NAME	进口道名称
	DIR	方向
	LANE_ID	车道 id
Flow.csv	ARM_ID	进口道 id
	START_TIME	统计开始时间
	END_TIME	统计结束时间
	VOLUMN_5MIN	5 分钟流量

点击下载

数据集下载

注：

（1）放行车道号“\_”前后分别代表进口道和车道，1, 2, 3, 4进口道分别对应方向为西，北，东，南。

（2）车道功能11, 12, 13分别代表直行，左转，右转。

【复赛】

待定

五、评估指标

【初赛】

评估指标：对称平均绝对百分比误差  $SMAPE$

$$SMAPE = \frac{1}{n} \sum_{t=1}^n \frac{|F_t - A_t|}{(|A_t| + |F_t|)/2 + \epsilon p}$$

式中  $A_t$  为第  $t$  个数据的真实值， $F_t$  为第  $t$  个数据的预测值， $n$  为样本总数， $\epsilon p$  是很小的数。

【复赛】

待定

六、提交说明

【初赛】

选手需要根据给定的训练集，预测出2023年8月28日的目标变量值（路口交通饱和度），并以csv格式（列形式）保存，文件命名为result.csv。

提交样例如下：

```
ID, START_TIME, END_TIME, SATURATION
1, 2023/8/28 0:00, 2023/8/28 0:05, 0.086638338
2, 2023/8/28 0:05, 2023/8/28 0:10, 0.015070817
3, 2023/8/28 0:10, 2023/8/28 0:15, 0.108573495
4, 2023/8/28 0:15, 2023/8/28 0:20, 0.108471395
5, 2023/8/28 0:20, 2023/8/28 0:25, 0.063634827
6, 2023/8/28 0:25, 2023/8/28 0:30, 0.076113913
7, 2023/8/28 0:30, 2023/8/28 0:35, 0.067542564
8, 2023/8/28 0:35, 2023/8/28 0:40, 0.021896137
```

注：选手需保证提交的result.csv中的预测数据的字段顺序与给定的提交样例顺序一致，提交的文件记录数应与提交示例中的记录数保持一致。为避免评测过程中文件编码格式导致评测错误的影响，文件统一以utf-8编码。

【初赛复核】

初赛复核阶段需要上传完整的、可复现的代码，并提供代码复现说明文档，在requirements.txt文件写明代码所使用的依赖包文件。

注：

（1）初赛复核阶段的代码提交格式详见**代码提交格式说明**。

（2）初赛阶段结束后，通过实名认证的TOP50团队晋级入围初赛评审环节，要求前50名团队提交最优成绩对应的训练预测代码进行审核，代码提交截止时间11月15日。组委会将审核并剔除存在作弊行为的队伍，晋级空缺名额递补。最终符合要求的排名前32名参赛队伍将进入复赛。

【复赛-A榜】

待定

【复赛-B榜】

待定

代码提交格式说明

代码提交规范目录结构如下。（需严格按照代码规范上传，否则导致的问题将由选手自行承担）

```
|-- project

    |-- code

    |-- train

    |-- model

    |-- raw_data

    |-- temp_data

    |-- result

        |-- result.csv

        |-- requirements.txt

        |-- description

            |-- description.md
```

1) 测试代码文件夹 code/

选手需要将针对测试集的测试代码放入该文件夹中，并提供main.py文件作为程序入口，确保可以通过执行该文件来运行训练、预测程序，得到最终结果。

2) 训练代码文件夹 train/

选手需要将针对训练集的训练代码放入该文件夹中，并在main.py中调用，确保可复现训练过程。

3) 模型文件夹 model/

选手需要将训练代码生成的模型文件放入该文件夹中，并注意不要将本地训练的模型文件上传到该文件夹。

4) 原始数据文件夹 raw\_data/

选手需将原始竞赛数据文件放入该文件夹。

5) 中间数据文件夹 temp\_data/

选手需要将必要的中间数据及额外数据放入该文件夹中。

6) 预测结果文件夹result/

选手需要将测评结果保存成.csv格式，并放入该文件夹。.csv文件中包含的内容可参考提交说明。

7) 依赖说明文件requirements.txt

选手需要将训练与测试代码所使用的依赖（需要安装的Python package）在requirements.txt文件中写明。

8) 说明文档文件夹description/

选手需要将含有算法思路、亮点解读、建模能力与环境的说明文档，以markdown格式进行保存，并放入该文件夹。说明文档内容可参考以下结构：

编程思路：简要说明算法编程实现方法和路径，提供核心代码段。

特色亮点：简要说明算法和程序在运算速度、效率等方面的特点和优势。

编程语言：采用的主要编程语言（必须注明版本号，如python3.7）。

软件运行环境：说明算法程序运行需配置的操作系统、语言环境、依赖包等（注明版本号）。

硬件配置环境：说明算法程序运行建议硬件环境配置。

运行说明：算法程序运行安装、配置、运行操作（必须）、算法结果等简要

