

“华为云杯”2020深圳开放数据应用创新大赛·深圳北站周边交通拥堵指数预测

AI大赛

“华为云杯”2020深圳开放数据应用创新大赛，大赛以“数聚粤港澳，智汇大湾区”为主题，面向全球征集基于开放数据的创新应用解决方案及优秀算法代码。

奖金

团队数

报名人数

大赛截止时间

¥ 180000

198

596

2020/06/19

举办方：深圳市政务服务数据管理局、南山区人民政府主办，南山区政务服务数据管理局、华为技术有限公司

大赛介绍

赛题说明

实时排名

求助交流



初赛报名时间：6月3日18点前

提交作品开放时间：5月5日-6月19日

关于大赛的更多信息，可前往大赛官网查阅：<https://www.sodic.com.cn/>

【赛题背景】

深圳北站作为深圳市的大型交通枢纽，其周边交通运行的好坏直接决定到达深圳北站的通行效率；而深圳路、南坪快速路等快速路，又是市区内的主要运输通道，对深圳北站的出行造成了直接影响。本赛题以此为背景，期望分析出深圳北站周边道路上，车辆对交通拥堵和北站出行的影响，实现基于网约车段交通拥堵指数的预测。

- 【更新汇总】
- 1、2020年4月25日，更新了276737、276738路段描述信息。

2、2020年4月25日，新增：可利用地图匹配GPS信息。

3、2020年4月25日，数据下载链接分成了4个，降低数据下载中断带来的损失。

4、2020年4月27日，2019年12月21日的数据，不计入评分成绩。

5、2020年4月27日，轨迹数据gps是GCJ02坐标系。

6、2020年6月02日，添加数据说明，用户ID是指“司机ID”。

7、2020年6月02日，更正数据说明，网约车轨迹数据第1列是订单id、第2列是用户id。

【赛题说明】

本赛题任务是对指定监测点在指定时刻的交通拥堵指数进行预测，要求选手在本地训练模型、预测并输出到csv文件中，格式如下：

```
id_sample, TTI
0,0.142857
1,0.285714
2,0.428571
.....
```

- id_sample：指待预测的“路段和时间”对应的编号，对应关系请见下载文件to_predict.csv。

• TTI：指交通拥堵指数。

【数据说明】

本赛题提供的训练集包括两个时间段，分别为从2019年1月到2019年3月，以及从2019年10月到2019年12限定在深圳北站周边，监测12个路段的交通拥堵指数。训练集的数据由两类数据构成：交通拥堵指数数据，和网约车轨迹数据。此赛题数据集可点击以下链接下载数据集二、数据集三、数据集四。测试集数据与训练集类似，不同点见后文。

1. 交通拥堵指数数据（train_TTI.csv）

交通拥堵指数的计算时间粒度为10分钟，即每个拥堵指数的数值表示的是某路段连续10分钟的平均拥堵指数段，分别是id_road、time、TTI和speed，字段说明如下

TTI [Ⓜ]	浮点 [Ⓜ]	交通拥堵指数traffic time indices的缩写 [Ⓜ]
speed [Ⓜ]	浮点 [Ⓜ]	平均车速 [Ⓜ]
time [Ⓜ]	时间类型 [Ⓜ]	此交通拥堵指数信息被记录的时间 [Ⓜ]

数据样例如下：

```
id_road,TTI,speed,timeⓂ
276183,1.11295,37.2711,2019-01-01 00:00:00Ⓜ
276183,1.05603,39.1086,2019-01-01 00:10:00Ⓜ
.....Ⓜ
```

路段范围和编号如下：

注：2020年4月25日，更新了276737、276738路段描述信息

```
276183, 南坪快速路：福龙-南坪立交到南坪-玉龙立交路段，西往东Ⓜ
276184, 南坪快速路：南坪-玉龙立交到福龙-南坪立交路段，东往西Ⓜ
275911, 福龙路：福龙-南坪立交到福龙-留仙大道立交路段，南往北Ⓜ
275912, 福龙路：福龙-留仙大道立交到福龙-南坪立交路段，北往南Ⓜ
276240, 留仙大道：福龙-留仙大道立交到留仙大道桥路段，西往东Ⓜ
276241, 留仙大道：留仙大道桥到福龙-留仙大道立交路段，东往西Ⓜ
276264, 玉龙路：南坪-玉龙立交到玉龙致远路口路段，南往北Ⓜ
276265, 玉龙路：玉龙致远路口到南坪-玉龙立交路段，北往南Ⓜ
276268, 新区大道：新区大道-民丰路口到新区大道-民宝路口路段，南往北Ⓜ
276269, 新区大道：新区大道-民宝路口到新区大道-白石龙路口路段，北往南Ⓜ
276737, 致远中路：致远-留仙路口到致远-民丰路口路段，北往南Ⓜ
276738, 致远中路：致远-民丰路口到致远-留仙路口路段，南往北Ⓜ
```

2. 网约车轨迹数据（train_201901_201903.zip, train_201910_201911.zip, train_20191201_20191220.zip）
此数据提供的是网约车的gps记录，这些gps记录按订单分组，每个订单都与上述路段相关。数据解压后为段，分别为订单id、用户id以及轨迹记录，轨迹记录是一个复合字段，包含多组gps记录，每组gps记录包是经纬度、速度、方向和时戳，字段说明如下（字段名不包含在数据中，仅在此处说明）。

注：可利用地图匹配GPS信息

字段名	类型	说明
id_order	加密字段	加密后的订单编号
id_user	加密字段	加密后的用户编号，此处指司机编号
gps_records	组合类型	5个字段为一组，[lng,lat,speed,direction,time]，分别是传感器记录的gps经度、纬度、速度、方向，以及该记录的时间。

数据样例如下：

```
48b...cd17,4d8...f21c7,"[114.01 22.58 20.97 99.09 1577855928, 114.02 22.58 28.97 99.10
1577855938, 114.02 22.59 20.97 99.09 1577855958,.....]"Ⓜ
53o...34xx,33b...f23s,"[112.01 21.58 40.97 98.09 1577855421, 112.02 21.58 58.22 98.10
1577855441, 112.02 22.59 50.97 98.09 1577855451,.....]"Ⓜ
.....Ⓜ
```

3. 测试集数据（toPredict_train_TTI.csv, toPredict_train_GPS.csv, toPredict_noLabel.csv）
测试集数据与训练集数据在字段上一致，与训练集的不同点在于“时间”和“标签”（即交通拥堵指数TTI）。时间和可供使用的数据。

对预测时间的限制，以下表样本数据为例（见toPredict_noLabel.csv），规定了待预测的信息。

```
id_sample,id_road,timeⓂ
0,276183,2019/12/22 08:30:00Ⓜ
.....Ⓜ
1,276184,2019/12/30 09:10:00Ⓜ
.....Ⓜ
```

对于此：

- 对于第1条记录，路段编号为276183，预测时间段为2019/12/22 08:30:00~2019/12/22 08:40:00这十分钟
- 对于第2条记录，路段编号为276184，预测时间段为2019/12/30 09:10:00~2019/12/30 09:20:00这十分钟

对使用数据的限制，禁止以任何形式利用未来的数据预测过去的的数据。例如，在预测2019/12/22 08:30:00~08:40:00的拥堵指数时，禁止使用2019/12/22 08:40:00这个时刻以后的数据（包括toPredict_train_TTI.csv toPredict_train_GPS.csv中的数据）。

待预测的和可利用数据的时间表如下：

注：2019年12月21日的数据，跟训练集有重合，故不计入评分成绩。

日期 [Ⓜ]	待预测时间段 [Ⓜ]	2019/12/21~2020/01/01期间可利用数据的时间段 [Ⓜ]
	toPredict_noLabel [Ⓜ]	toPredict_train_TTI, toPredict_train_GPS [Ⓜ]
2019/12/21 [Ⓜ]	[08:30, 09:00) [Ⓜ]	当日08:30以前的数据（当日7:30以前的数据未提供） [Ⓜ]
2019/12/22 [Ⓜ]	[10:30, 11:00) [Ⓜ]	当日10:30以前的数据（08:30~09:30之间的数据未提供） [Ⓜ]
2019/12/23 [Ⓜ]	[12:30, 13:00) [Ⓜ]	当日12:30以前的数据（10:30~11:30之间的数据未提供） [Ⓜ]
2019/12/24 [Ⓜ]	[14:30, 15:00) [Ⓜ]	当日14:30以前的数据（12:30~13:30之间的数据未提供） [Ⓜ]
2019/12/25 [Ⓜ]	[16:30, 17:00) [Ⓜ]	当日16:30以前的数据（14:30~15:30之间的数据未提供） [Ⓜ]
2019/12/26 [Ⓜ]	[18:30, 19:00) [Ⓜ]	当日18:30以前的数据（16:30~17:30之间的数据未提供） [Ⓜ]
[Ⓜ]	[20:30, 21:00) [Ⓜ]	当日20:30以前的数据（18:30~19:30之间的数据未提供） [Ⓜ]
2019/12/27 [Ⓜ]	[09:00, 09:30) [Ⓜ]	当日09:00以前的数据（当日8:00以前的数据未提供） [Ⓜ]
2019/12/28 [Ⓜ]	[11:00, 11:30) [Ⓜ]	当日11:00以前的数据（09:00~10:00之间的数据未提供） [Ⓜ]
2019/12/29 [Ⓜ]	[13:00, 13:30) [Ⓜ]	当日13:00以前的数据（11:00~12:00之间的数据未提供） [Ⓜ]
2019/12/30 [Ⓜ]	[15:00, 15:30) [Ⓜ]	当日15:00以前的数据（13:00~14:00之间的数据未提供） [Ⓜ]
2019/12/31 [Ⓜ]	[17:00, 17:30) [Ⓜ]	当日17:00以前的数据（15:00~16:00之间的数据未提供） [Ⓜ]
2020/01/01 [Ⓜ]	[19:00, 19:30) [Ⓜ]	当日19:00以前的数据（17:00~18:00之间的数据未提供） [Ⓜ]
	[21:00, 21:30) [Ⓜ]	当日21:00以前的数据（19:00~20:00之间的数据未提供） [Ⓜ]

【评分标准】

初赛和决赛均采用mae（mean absolute error）作为评分指标，计算公式如下：

$$mae = \frac{1}{n} \sum_k^n |y_k - \hat{y}_k|$$

其中，k为样本编号，n为样本总数量，y_k为第k个样本的真实值，(y_k)[^]为第k个样本的预测值。
mae的计算脚本可参数如下代码：

```
# python代码
from sklearn.metrics import mean_absolute_error
def calc_mae(y_true, y_pred):
    return mean_absolute_error(y_true, y_pred)

# 调用函数计算mae值
y_true = [1,2,3,4]
y_pred = [1.1,1.2,1.3,1.4]
print(calc_mae(y_true,y_pred)) # 输入结果为1.3
```

【模型规范】

1) 推荐选手采用python 3.6以上的版本，以便审核代码。

【提交说明】

所有参赛者需下载数据，在本地训练并得到预测结果；然后登陆华为云大赛平台提交结果文件，提交窗口地址：<https://competition.huaweicloud.com/information/1000040088/introduction?track=107>（提交地址：<https://competition.huaweicloud.com/information/1000040088/introduction?track=107>）
数据文件格式要求请见“赛题说明”，亦在此赛题提供的数据压缩包中找到。

提交方法

请参考数据文件中的submit_example.csv的格式进行提交，文件含两列：

```
id_sample, TTI
0,0.142857
1,0.285714
2,0.428571
.....
```

评分说明

- 1) 本次比赛初赛时间为：5月5日10:00-6月19日14:00。
决赛提交时间段为：7月1日10:00- 7月10日10:00。
(初赛作品提交时间及评分说明，5月31日变更，请选手们留意查看时间点)
- 2) 每个团队每天有3次评测机会，所提交的模型得分可在大赛平台页面“提交作品”中查询。
- 3) 排行榜每6个小时刷新一次。

参考资料



7*24
多渠道服务支持

0元
免费备案

专业服务
云业务全流程支持

退订
享无忧退订服务

建议反馈
优化改进建议

热门产品

云服务器 云速邮箱 域名注册 云速建站

📄 下载华为云APP

关注我们



售前咨询：950808

法律声明 | 隐私政策

版权所有 © 华为软件技术有限公司 2020 保留一切权利 苏ICP备17040376号-32 苏B2-20130048号 代理域名注册服务机构：新网
苏公网安备 32011402010636号