

数据集及任务描述

数据集:

[https://tianchi.aliyun.com/competition/information.htm?
spm=5176.100067.5678.2.jIKgwp&raceId=231597](https://tianchi.aliyun.com/competition/information.htm?spm=5176.100067.5678.2.jIKgwp&raceId=231597)

问题缘起:

高速路口是众所周知的交通瓶颈，交通拥堵的时候高速路口排成的长龙实在是可怕。只有能够准确的预测交通流量一些应对措施才行得通。例如如果能够预测到一小时后会有拥堵，那么交通主管部门可以立即部署增加收费员或者在上流路口分流交通。交通流受很多复杂因素影响比如天气，假期，星期几，几点，预测交通流及ETA(预计到底时间)是一个已知的挑战。

如果本次大赛选手能够准确的预测交通流和ETA可以帮助交通部门利用大数据舒缓高速路口的交通状况。

任务

任务1: 估计从指定路口到高速路口的平均耗时

为每20分钟的时间段估计特定路线的平均通过时间:

- a路线: Intersection A to Tollgates 2 & 3;
- b路线: Intersection B to Tollgates 1 & 3;
- c路线: Intersection C to Tollgates 1 & 3.

提交文件如表1

| Field | Type | Description |
|-----------------|--------|---|
| intersection_id | string | intersection ID |
| tollgate_id | string | tollgate ID |
| time_window | string | e.g., [2016-09-18 08:40:00,2016-09-18 09:00:00) |
| avg_travel_time | float | average travel time (seconds) |

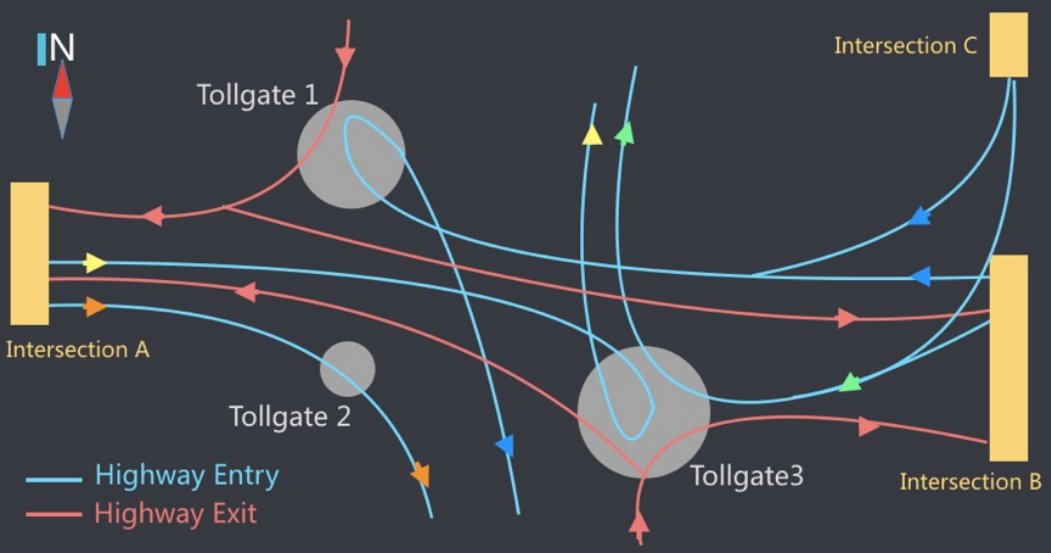
任务1评价标准 *Mean Absolute Percentage Error (MAPE)*

任务2: 预测高速路口平均交通流量

预测每20分钟时间段 tollgates 1, 2 and 3的进出车辆流量，注意 tollgate 2只允许从这里进高速公路，其他两个可进可出 提交文件如表2

| <i>Field</i> | <i>Type</i> | <i>Description</i> |
|--------------------|-------------|---|
| <i>tollgate_id</i> | string | tollgate ID |
| <i>time_window</i> | string | e.g., [2016-09-18 08:40:00,2016-09-18 09:00:00) |
| <i>direction</i> | string | 0: entry, 1: exit |
| <i>volume</i> | int | total volume |

图1目标区域路网图



测试集与训练集

目标区域路网topology (Tables 3 and 4), 车辆轨道线 (Table 5), 路口历史交通流量 (Table 6),天气(Table 7).

测试集

比赛初期使用的测试集为10月18日到24日的数据，如图，红色的时间段为需要选手预测的交通时间段，绿色的为给定的。需要预测的为 08:00 - 10:00 以及17:00 - 19:00 内的20分钟时间段。



在3月25日有一次数据交换测试集换成10月25日至10月31日的交通数据

训练集

对于Table 5包含7月19日至 10月17日，Table 6初始数据集包含9月19日至10月17日；3月25日数据交换后两个数据均增加10月18日到24日的数据

数据集描述

表3为link特征

| <i>Field</i> | <i>Type</i> | <i>Description</i> |
|--------------------------|--------------------|--|
| <i>link_id</i> | string | link id |
| <i>length</i> | float | length (meter) |
| <i>width</i> | float | length (meter) |
| <i>lanes</i> | int | number of lanes |
| <i>in_top</i> | string | incoming road link(s), separated by comma (as shown in Figure 3) |
| <i>out_top</i> | string | outgoing road link(s), separated by comma (as shown in Figure 3) |
| <i>lane_width</i> | float | lane width (meter) |

如图一条路线由一系列link组成

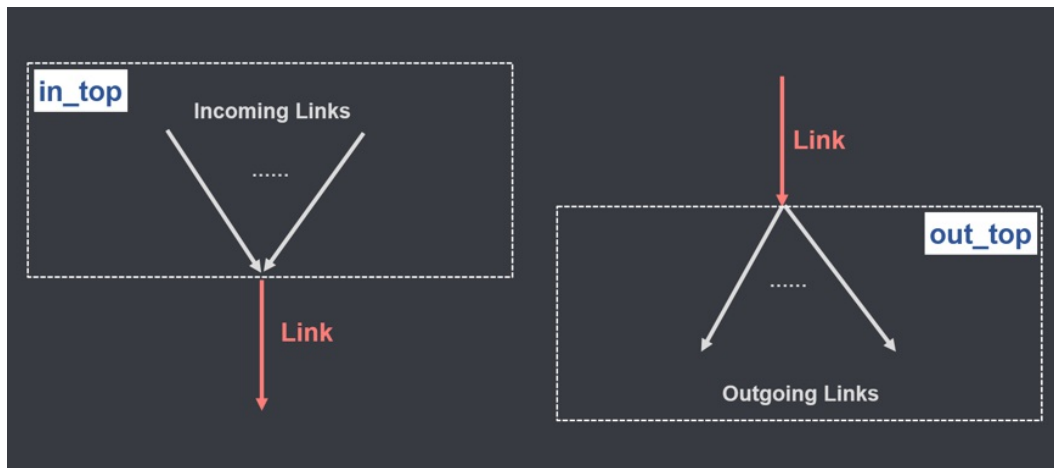


表4为每条路线组成的link

| <i>Field</i> | <i>Type</i> | <i>Description</i> |
|------------------------|-------------|--|
| <i>intersection_id</i> | string | intersection ID |
| <i>tollgate_id</i> | string | tollgate ID |
| <i>link_seq</i> | string | a sequence of link IDs from the intersection to the tollgate separated by commas (shown in Figure 4) |

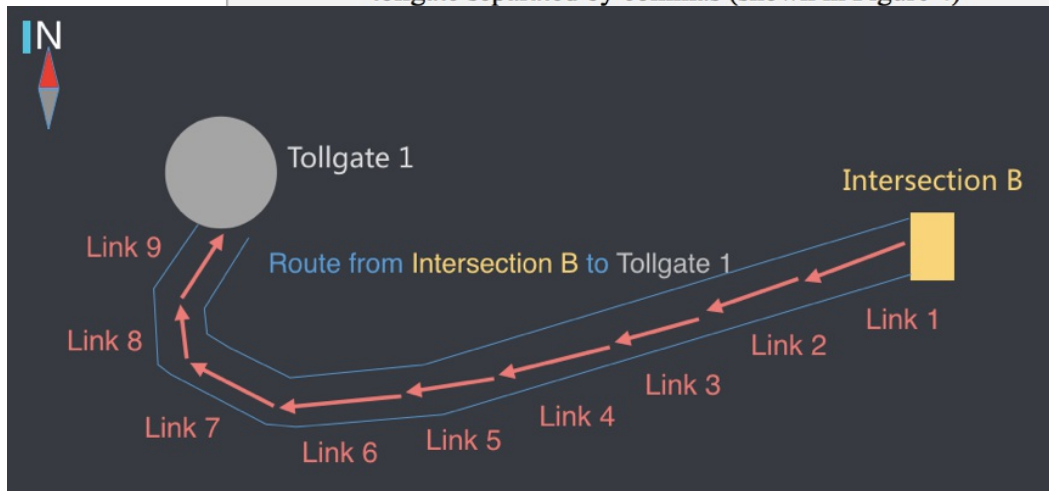


表5为实际车辆沿着路线从路口到高速路口的时间戳

| <i>Field</i> | <i>Type</i> | <i>Description</i> |
|------------------------|-------------|--|
| <i>intersection_id</i> | string | intersection ID |
| <i>tollgate_id</i> | string | tollgate ID |
| <i>vehicle_id</i> | string | vehicle ID |
| <i>starting_time</i> | datetime | time point when the vehicle enters the route |
| <i>travel_seq</i> | string | trajectory in the form of a sequence of link traces separated by ";", each trace consists of link id, enter time, and travel time in seconds, separated by "#" |
| <i>travel_time</i> | float | the total time (in seconds) that the vehicle takes to travel from the intersection to the tollgate |

表6为通过高速路口收费站的车流

| <i>Field</i> | <i>Type</i> | <i>Description</i> |
|----------------------|-------------|---|
| <i>time</i> | datetime | the time when a vehicle passes the tollgate |
| <i>tollgate_id</i> | string | ID of the tollgate |
| <i>direction</i> | string | 0: entry, 1: exit |
| <i>vehicle_model</i> | int | this number ranges from 0 to 7, which indicates the capacity of the vehicle (bigger the higher) |
| <i>has_etc</i> | string | does the vehicle use ETC (Electronic Toll Collection) device? 0: No, 1: Yes |
| <i>vehicle_type</i> | string | vehicle type: 0-passenger vehicle, 1-cargo vehicle |

表7为目标区域天气状况

| <i>Field</i> | <i>Type</i> | <i>Description</i> |
|-----------------------|-------------|--------------------------------------|
| <i>date</i> | date | date |
| <i>hour</i> | int | hour |
| <i>pressure</i> | float | air pressure (hPa: Hundred Pa) |
| <i>sea_pressure</i> | float | sea level pressure (hPa: Hundred Pa) |
| <i>wind_direction</i> | float | wind direction (°) |
| <i>wind_speed</i> | float | wind speed (m/s) |
| <i>temperature</i> | float | temperature (°C) |
| <i>rel_humidity</i> | float | relative humidity |
| <i>precipitation</i> | float | precipitation (mm) |

评价标准

Mean Absolute Percentage Error (MAPE)

任务1

d_{rt} 和 p_{rt} 分别为路线 r , 时间段 t 上实际与预测的到达高速路口所需时间, R, T 为路线数和需要预测的时间段数

$$MAPE = \frac{1}{R} \sum_{r=1}^R \left(\frac{1}{T} \sum_{t=1}^T \left| \frac{d_{rt} - p_{rt}}{d_{rt}} \right| \right)$$

任务2

C 为高速路口-进出组合数 $pairs$ (有5种组合: 1-entry, 1-exit, 2-entry, 3-entry and 3-exit), T 需要预测的时间段数需要预测的时间段数, f_{ct} 和 p_{ct} 为实际及预测的该时间段交通流量

$$MAPE = \frac{1}{C} \sum_{c=1}^C \left(\frac{1}{T} \sum_{t=1}^T \left| \frac{f_{ct} - p_{ct}}{f_{ct}} \right| \right)$$