

# 全球航运轨迹预测

## 一、赛题背景

全球贸易飞速发展的当下，水上交通占据主导地位。不同于有固定道路的路上交通，水上交通更为复杂，路径规划更具挑战性。为提升船舶水上交通的决策水平，准确生成船舶的轨迹至关重要。本次挑战赛旨在通过分析船舶 AIS 信息数据，较为准确地生成在途船舶的剩余航行轨迹，从而提升水上交通的管理水平。

## 二、赛题任务

参赛者需要利用提供的船舶 AIS 信息数据集，构建模型预测在途船舶的剩余轨迹（一系列 `longitude` 和 `latitude`）。选手需要自行进行数据预处理操作，应当充分利用提供数据集中各个字段的特征，并搭建算法模型来进行在途船舶的剩余轨迹生成，并要求模型可以自主判断是否停止生成，从而达到模拟到港的效果。

- 鼓励参赛者运用先进的数据分析、机器学习、深度学习等技术。
- 允许使用额外的公开数据源，以增强模型的准确性和泛化能力。
- 重视数据预处理和特征工程，以提高模型性能。

## 三、数据描述

1. 船舶 AIS 信息数据训练集：包含完整的历史 AIS 船舶从离港到到港，在途航行的经纬度、状态、吃水、船艏向、航迹向、对地速度等关键参数。

2.船舶 AIS 信息数据测试集：仅包含部分航段的 AIS 信息，包括在途航行的经纬度、状态、吃水、船艏向、航迹向、对地速度。

3.港口静态信息数据：包含有关港口的基本描述，如港口代码、名称、位置、国家、类型等信息。

#### 四、模型预测结果评分标准

选手需要生成在途船舶的剩余轨迹并使模型自主停止生成，以模拟到港。当模型生成的剩余轨迹长度为 0 时，该次轨迹生成的得分将是 0 分。当模型生成的剩余轨迹长度大于等于 1 时，将继续比较选手生成的轨迹和真实轨迹之间的差距，匹配与选手生成的轨迹点的对应时间最接近的真实点，并计算匹配后的生成轨迹点和真实点之间的距离之和  $D1$ 。若选手生成的轨迹点数与真实的轨迹点数不匹配，则未匹配上的点数将与经纬度都为 0 的虚拟点计算距离并求和  $D2$ 。 $100/(1+D1+D2)$ 即为该条轨迹生成的分数  $F$ ，满分为 100 分。

假设测试集中有  $n$  条航段轨迹，则最终的评分将为

$$\frac{\sum F}{n}$$

其中， $D1$ ， $D2$ ， $F$  的具体计算公式如下：

$$D1 = \sum \sqrt{(\text{long}_{\text{true},i} - \text{long}_{\text{pred},i})^2 + (\text{lat}_{\text{true},i} - \text{lat}_{\text{pred},i})^2}$$

$$D2 = \sum \sqrt{\text{long}_{\text{mismatch}}^2 + \text{lat}_{\text{mismatch}}^2}$$

$$F = \frac{100}{1 + D1 + D2}$$

最终模型预测结果评分满分为 100 分。

#### 五、提交要求

参赛队伍需提交一个包含模型代码和预测结果的压缩文件。预测结果需包括每条航段的给定轨迹和特征参数，以及生成的轨迹。

## 六、数据格式

船舶 AIS 信息数据训练集共有 100 条船舶的 295810 条 AIS 信息数据（ship\_name 在三张船舶 AIS 信息表中贯穿一致），数据说明如下：

列名	类型	说明
ship_name	int	船舶的名称（已用阿拉伯数字代替）
slice_time	str	AIS 当前时间对应的切片时间，格式为 YYYY-MM-dd HH:MM:SS+tz
longitude	float	AIS 航段当前位置经度
latitude	float	AIS 航段当前位置纬度
status	str	AIS 航段当前位置对应的船舶状态（5 代表靠泊，1 代表锚泊，0 代表航行）
hdg	float	船艏向，船艏朝向的方向
cog	float	航迹向，船舶航行的方向
sog	float	对地速度，船舶的对地速度
draught	float	AIS 航段当前位置的吃水

船舶 AIS 信息数据测试集共有 97 条船舶的 59560 条 AIS 信息数据，slice\_time 为三小时一个点位，选手需要自行判断航段划分，数据说明如下：

列名	类型	说明
ship_name	int	船舶的名称（已用阿拉伯数字代替）
slice_time	str	AIS 当前时间对应的切片时间，格式为 YYYY-MM-dd HH:MM:SS+tz

longitude	float	AIS 航段当前位置经度
latitude	float	AIS 航段当前位置纬度
status	str	AIS 航段当前位置对应的船舶状态（5 代表靠泊，1 代表锚泊，0 代表航行）
hdg	float	船艏向，船艏朝向的方向
cog	float	航迹向，船舶航行的方向
sog	float	对地速度，船舶的对地速度
draught	float	AIS 航段当前位置的吃水

船舶 AIS 信息数据测试集\_真值共有 97 条船舶的 58765 条 AIS 信息数据，该数据集是船舶 AIS 信息数据测试集的后续航程轨迹，数据说明如下：

列名	类型	说明
ship_name	int	船舶的名称（已用阿拉伯数字代替）
slice_time	str	AIS 当前时间对应的切片时间，格式为 YYYY-MM-dd HH:MM:SS+tz
longitude	float	AIS 航段当前位置经度
latitude	float	AIS 航段当前位置纬度
status	str	AIS 航段当前位置对应的船舶状态（5 代表靠泊，1 代表锚泊，0 代表航行）
hdg	float	船艏向，船艏朝向的方向
cog	float	航迹向，船舶航行的方向
sog	float	对地速度，船舶的对地速度
draught	float	AIS 航段当前位置的吃水

leg_end_port_code	str	AIS 当前航段的目的地
-------------------	-----	--------------

港口静态信息数据包含 154 条港口的静态信息，数据格式说明如下：

列名	类型	说明
port_code	str	港口代码（字符串类型）
ctry_code	str	国家代码（字符串类型）
name_en	str	英文名称（字符串类型）
name_cn	str	中文名称（字符串类型）
lon	float	经度（浮点数类型）
lat	float	纬度（浮点数类型）
timezone_offset	int	时区偏移（整数类型，单位为小时）