

贝贝人工智能挑战赛：订单物流时效预测

赛题描述

对于各大电商平台而言，良好的物流体验是提升用户满意度的一个重要手段。而准确预估订单的送达时间，不仅可以为用户下单提供有效决策，也可以让用户提前感知，合理规划收货时间。我们把如何准确预估每笔订单的送达时间，定义为订单物流时效预测问题。

该问题的主要挑战是整个订单的履约流程较长，不仅涉及到下单-发货-揽件-走件-签收等多个环节，而且每个环节影响的因素也截然不同。同时，由于订单天然存在密度差异性，因此无论是从商家还是物流公司还是消费者（收货地）角度看，数据都不可避免地存在不平衡问题。因此本次大赛需要选手综合运用数学建模，模型设计，数据挖掘等多种技术手段，从而构建出高精度的算法模型，提升订单物流时效预测的精度。

组队规则

参赛队伍可1-3人组队参赛，竞赛报名以及组队队员变更截止时间为2019年10月7日22点。

赛程安排

初赛

1. 平台会在2019年9月18日前开放训练数据集A、测试集A，参赛选手可以自行下载数据，在本地进行算法设计、模型训练及评估。[下载链接](#) 提取码: 6v1h
2. 参赛选手在本地使用先前训练的模型进行预测，生成预测结果并提交至平台。结果提交后，系统会按照评测指标实时反馈分数，并更新榜单排名。
3. 每队每天最多可提交1次。榜单以所有参赛队伍的历史最优成绩进行排名。当有团队提交新的预测结果之后，榜单将实时更新客观指标。
4. 初赛提交截止时间为2019年10月7日22点，参赛队伍需要在截止时间前提交最终版代码与初赛报告到大赛公邮eedcup@dian.org.cn

注：大赛主办方可以根据实际时间安排，对赛程进行更新。

数据说明

本次比赛我们分别为初赛和复赛（复赛数据量更大）各提供了一份真实场景下的数据集，每份数据集都包括训练集、测试集两个部分。训练、测试数据集的格式如下，所有字段间用Tab键（`\t`）分隔，Utf-8编码。以下标注*的字段为测试集时不可用字段。以上数据进行过按比例抽样和脱敏操作。

字段	数据类型	备注
uid 买家id	int	
plat_form 交易平台	int	贝贝/贝店/贝仓...
biz_type 业务来源	int	不同业务线售卖
create_time 订单创建时间	string yyyy-MM-dd hh:mm:ss	
payed_time 订单支付时间	string yyyy-MM-dd hh:mm:ss	
product_id 商品id	int	
cate1_id 商品类目1级	int	
cate2_id 商品类目2级	int	
cate3_id 商品类目3级	int	
preselling_shipped_time 预售发货时间	string yyyy-MM-dd hh:mm:ss	存在噪声数据
seller_uid 商家id	int	
company_name 商家公司id	int	
*lgst_company 物流公司id	int	
*warehouse_id 仓库id	int	
*shipped_prov_id 发货省份id	int	
*shipped_city_id 发货城市id	int	
rvcr_prov_name 收货省份id	int	
rvcr_city_name 收货城市id	int	
*shipped_time 发货时间	string yyyy-MM-dd hh:mm:ss	取决于商家填写（可能存在延迟）
*got_time 揽件时间	string yyyy-MM-dd hh:mm:ss	取决于物流公司信息（可能存在延迟）
*dlved_time 走件时间	string yyyy-MM-dd hh:mm:ss	
*signed_time 签收时间	string yyyy-MM-dd hh:mm:ss	

评价标准

输出：日期 时刻 (使用Utf-8编码)

eg:

```
2019-05-04 8
2019-05-05 7
2019-05-06 6
2019-05-07 5
2019-05-08 4
```

误差率 rankScore = MSE(real_signed_time ,pred_signed_time) 精确到小时

准时率 onTimePercent = Count(pred_singed_date <= real_signed_date) 精确到天

解释：本次评分考虑两个因子，一方面考虑到货时间均方根误差，另外一个就是考虑是否有天级别延迟。排名首先要求准时率>=98%,在此基础上对误差率进行排名。

注：2019种子杯参赛交流群：752849984