


	状态	举办方	第 1 赛季截止日期	总奖池	参赛队
IJCAI-18 阿里妈妈搜索...	进行中	 阿里妈妈  TIANCHI 天池	2018/04/23	\$37000	3410

- 报名参赛
- 赛制介绍 (./introduction.htm?)
- 赛题与数据 (./information.htm?)
- FAQ (./faq.htm?)
- 排行榜  (./rankingList.htm?)
- 技术圈

文件名称 (报名后可下载)	文件格式
round1_ijcai_18_result_demo_20180301.txt	.txt (147B)
round1_ijcai_18_test_a_20180301.zip	.zip (1MB)
round1_ijcai_18_train_20180301.zip	.zip (39MB)

赛题内容

本次比赛以阿里电商广告为研究对象，提供了淘宝平台的海量真实交易数据，参赛选手通过人工智能技术构建预测模型预估用户买意向，即给定广告点击相关的用户（user）、广告商品（ad）、检索词（query）、上下文内容（context）、商店（shop）等信息的条件下预测广告产生购买行为的概率（pCVR），形式化定义为：pCVR=P(conversion=1 | query, user, ad, context, shop)。

结合淘宝平台的业务场景和不同的流量特点，我们定义了以下两类挑战：

(1) 日常的转化率预估

(2) 特殊日期的转化率预估

评估指标

通过logarithmic loss（记为logloss）评估模型效果（越小越好），公式如下：

$$\log loss = -\frac{1}{N} \sum_{i=1}^N (y_i \log(p_i) + (1 - y_i) \log(1 - p_i))$$

其中N表示测试集样本数量，yi表示测试集中第i个样本的真实标签，pi表示第i个样本的预估转化率。

数据说明

本次比赛为参赛选手提供了5类数据（基础数据、广告商品信息、用户信息、上下文信息和店铺信息）。基础数据表提供了搜索广告最基本的信息，以及“是否交易”的标记。广告商品信息、用户信息、上下文信息和店铺信息等4类数据，提供了对转化率预估可能帮助的辅助信息。

用于初赛的数据包含了若干天的样本。最后一天的数据用于结果评测，对选手不公布；其余日期的数据作为训练数据，提供给参手；。

在上述各张数据表中，绝大部分样本包含了完整的字段数据，也有少部分样本缺乏特定字段的数据。如果一条样本的某个字段为“-1”，表示这个样本的对应字段缺乏数据。

3月1日补充说明：

关于样本和字段分隔符的说明：

每条样本单独占一行，使用“\r”（回车符）进行分割。同一条样本的各个字段，使用空格进行分割。冒号 “:”、逗号“,”和分号“;”都是字段内部的分隔符号。各个字段的含义和拼接格式，请参见“数据说明”部分。

使用 Windows 系统的选手请注意，使用Windows记事本打开比赛数据时无法正常显示换行格式，请使用 Notepad++ 或 UltraEdit 等文本编辑器进行浏览。

基础数据

字段	解释
instance_id	样本编号，Long
is_trade	是否交易的标记位，Int类型；取值是0或者1，其中1 表示这条样本最终产生交易，0 表示没有交易
item_id	广告商品编号，Long类型
user_id	用户的编号，Long类型
context_id	上下文信息的编号，Long类型
shop_id	店铺的编号，Long类型

广告商品信息

字段	解释
item_id	广告商品编号，Long类型
item_category_list	广告商品的的类目列表，String类型；从根类目（最粗略的一级类目）向叶子类目（最精细的类目）依次排列，数据拼接格式为 "category_0;category_1;category_2"，其中 category_1 是 category_0 的子类目，category_2 是 category_1 的子类目
item_property_list	广告商品的属性列表，String类型；数据拼接格式为 "property_0;property_1;property_2"，各个属性没有从属关系
item_brand_id	广告商品的品牌编号，Long类型
item_city_id	广告商品的城市编号，Long类型
item_price_level	广告商品的价格等级，Int类型；取值从0开始，数值越大表示价格越高
item_sales_level	广告商品的销量等级，Int类型；取值从0开始，数值越大表示销量越大
item_collected_level	广告商品被收藏次数的等级，Int类型；取值从0开始，数值越大表示被收藏次数越大
item_pv_level	广告商品被展示次数的等级，Int类型；取值从0开始，数值越大表示被展示次数越大

用户信息

字段	解释
user_id	用户的编号，Long类型
user_gender_id	用户的预测性别编号，Int类型；0表示女性用户，1表示男性用户，2表示家庭用户
user_age_level	用户的预测年龄等级，Int类型；数值越大表示年龄越大
user_occupation_id	用户的预测职业编号，Int类型
user_star_level	用户的星级编号，Int类型；数值越大表示用户的星级越高

上下文信息

字段	解释
context_id	上下文信息的编号，Long类型
context_timestamp	广告商品的展示时间，Long类型；取值是以秒为单位的Unix时间戳，以1天为单位对时间戳进行了偏移
context_page_id	广告商品的展示页面编号，Int类型；取值从1开始，依次增加；在一次搜索的展示结果中第一屏的编号为1，第二屏的编号为2
predict_category_property	根据查询词预测的类目属性列表，String类型；数据拼接格式为“category_A:property_A_1,property_A_2,property_A_3;category_B:-1;category_C:property_C_1,property_C_2”，其中 category_A、category_B、category_C 是预测的三个类目；property_B 取值为-1，表示预测的第二个类目 category_B 没有对应的预测属性

店铺信息

字段	解释
shop_id	店铺的编号，Long类型
shop_review_num_level	店铺的评价数量等级，Int类型；取值从0开始，数值越大表示评价数量越多
shop_review_positive_rate	店铺的好评率，Double类型；取值在0到1之间，数值越大表示好评率越高
shop_star_level	店铺的星级编号，Int类型；取值从0开始，数值越大表示店铺的星级越高
shop_score_service	店铺的服务态度评分，Double类型；取值在0到1之间，数值越大表示评分越高
shop_score_delivery	店铺的物流服务评分，Double类型；取值在0到1之间，数值越大表示评分越高
shop_score_description	店铺的描述相符评分，Double类型；取值在0到1之间，数值越大表示评分越高

关于我们 (/aboutUs.htm)

法律协议 (/userAgreement.htm)

了解更多，请

解决方案

行业架构师咨询

成熟解决方案 (https://www.aliyun.com/solution/all)

|

专业架构师咨询 (https://www.aliyun.com/solution/all)

|

成功客户案例分享 (https://www.aliyun.com/solution/all)

|

专属定制服务 (https://www.aliyun.com/solution/all)

热门产品

云服务器ECS
(https://www.aliyun.com/product/ecs?spm=5176.8142029.s0114624587142029.0.0#6.715c4636Lyprub)

云数据库RDS
(https://www.aliyun.com/product/rds?spm=5176.8142029.s0114624587142029.0.0#6.715c4636Lyprub)

云存储OSS
(https://www.aliyun.com/product/oss?spm=5176.8142029.s0114624587142029.0.0#6.715c4636Lyprub)

NAT网关
(https://www.aliyun.com/product/nat?spm=5176.8142029.s0114624587142029.0.0#6.715c4636Lyprub)

负载均衡
(https://www.aliyun.com/product/alb?spm=5176.8142029.s0114624587142029.0.0#6.715c4636Lyprub)

域名注册
(https://www.aliyun.com/product/aliyun.com/market.aliyun.com/)

网站建设
(https://www.aliyun.com/product/aliyun.com/market.aliyun.com/)

大数据
(https://www.aliyun.com/product/aliyun.com/market.aliyun.com/)

	网站备案 (https://beian.aliyun.com/)	网安法 (https://market.aliyun.com/security)	CDN加速 (https://www.aliyun.com/cdn)	API网关 (https://www.aliyun.com/apigateway)	企业邮箱 (https://www.aliyun.com/enterprise-mail)	whois查询 (https://whois.aliyun.com/)	视频直播 (https://www.aliyun.com/video-live)	视频转码 (https://www.aliyun.com/video-transcode)
更多推荐	全民云计算 (https://promotion.aliyun.com/qa/qitem.htm?spm=5176.8142029.631162.98.0IMSSV&recordcode=33010002000099)	免费套餐 (https://market.aliyun.com/special)	学生机 (https://promotion.aliyun.com/qa/qitem.htm?spm=5176.8142029.631162.98.0IMSSV&recordcode=33010002000099)	IT论坛 (https://bbs.aliyun.com/)	数据可视化 (https://data.aliyun.com/)	云虚拟机 (https://cloud.aliyun.com/)	com域名 (https://www.aliyun.com/domain/com)	cn域名 (https://www.aliyun.com/domain/cn)
	合规安全解决方案 (https://www.aliyun.com/solution/security/compliance)							

[关于我们 \(\)](#) [法律声明 \(\)](#) [廉正举报 \(\)](#) [友情链接 \(\)](#)

阿里巴巴集团 (<http://www.alibabagroup.com/cn/global/home>) 淘宝网 (<http://www.taobao.com/>) 天猫 (<http://www.tmall.com/>) 聚划算 (<http://ju.taobao.com/>) 全球速卖通 (<http://www.aliexpress.com/>)
阿里巴巴国际交易市场 (<http://www.alibaba.com/>) 1688 (<http://www.1688.com/>) 阿里妈妈 (<http://www.alimama.com/index.htm>) 飞猪 (<http://www.alitrip.com/>) 阿里云计算 (<http://www.aliyun.com>)
YunOS (<http://www.yunos.com/>) 阿里通信 (<http://aliqin.tmall.com/>) 万网 (<http://wanwang.aliyun.com/>) 高德 (<http://www.autonavi.com/>) UC (<http://www.uc.cn/>) 虾米 (<http://www.xiami.com>)
阿里星球 (<http://www.alibabaplanet.com>) 来往 (<http://www.laiwang.com/>) 钉钉 (<http://www.dingtalk.com/?lwfrom=20150205111943449>) 支付宝 (<https://www.alipay.com/>)

© 2009-2016 Aliyun.com 版权所有 ICP证：浙B2-20080101

 ([//verify.nic.xin/xinDetail/xinAuthInfoDetail?domainName=aliyun.xin](http://verify.nic.xin/xinDetail/xinAuthInfoDetail?domainName=aliyun.xin))  ()
 浙公网安备 33010002000099号 (<http://www.beian.gov.cn/portal/registerSystemInfo?spm=5176.8142029.631162.98.0IMSSV&recordcode=33010002000099>)