

# ICME 短视频竞赛 FAQ 汇总（2.18）

整理了竞赛讨论区、微信群、邮件等渠道反馈和收集的问题，汇总成 FAQ，仅供参考。

## 2 月 18 日更新

### 【数据集】

Q: 训练集、测试集是否按照行为时间划分？

A: 数据的先后顺序与时间有一定的相关性，可以根据时间的相对位置进行大致判断。

Q: 如何划定验证集？

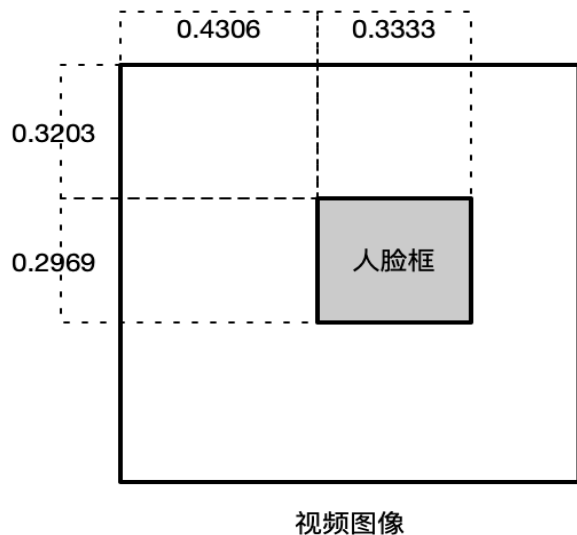
A: 建议按照数据先后顺序划定验证集。

Q: 有些字段的值为-1，是什么原因？

A: track1 中的字段 user\_city 为空，默认置为-1。若某条数据的某个字段为空，默认置为-1。

Q: 人脸特征相对位置的描述如何用图像示例？

A: relative\_position:[0.4306, 0.3203, 0.3333, 0.2969]，依次为左上角相对于视频宽高位置，以及人脸框的宽高相对于视频宽高位置。



### 【评估方法】

Q: 评估公式是什么？先计算 AUC 还是先加权预测结果？

A: 评估公式为： $0.7 * \text{finish} + 0.3 * \text{like}$ 。先计算两个指标的 AUC，再进行加权算总分。

**【组队与提交】**

Q: 组队邀请时, 对方无法接收到邮件。

A: 请将详细情况和截图发送至 [support@biendata.com](mailto:support@biendata.com), 会有服务商工作人员解决; 若长时间无人处理, 请在微信群@小助手。

Q: 提交结果时, 无法接收到验证码。

A: 请将详细情况和截图发送至 [support@biendata.com](mailto:support@biendata.com), 会有服务商工作人员解决; 若长时间无人处理, 请在微信群@小助手。

Q: 每日提交次数的刷新时间是北京时间几点?

A: 竞赛使用 UTC 国际协调时间, 每日 0 时对应北京时间早上 8 时。

Q: 提交时报错, 怎么办?

A: 提交文件格式示例, 在**【数据】**板块可以下载, 请严格按照上传要求进行操作。若仍有其他问题, 请将详细情况和截图发送至 [support@biendata.com](mailto:support@biendata.com), 会有服务商工作人员解决; 若长时间无人处理, 请在微信群@小助手。

Q: 多久更新一次榜单?

A: 每 10 分钟刷新一次榜单排名。

Q: 组队的截止时间是哪天?

A: 组队的截止时间是 4 月 1 日。

## 2月1日更新

Q: 想请问一下作品城市 (item\_city) 是什么含义?

A: 作品城市 (item\_city) 指的是发布作品时的城市信息。

Q:

1) 为什么 idx 需要添加位置偏移

2) idx 和 features 的区别除了添加了位置偏移, 就只是 int32 和 float32 的区别。

3) 在 FM 模型中,  $f_e_m = \text{tf.multiply}(\text{feature\_values}, \text{embeddings})$ , 为什么 embeddings 还需要乘上 feature\_values?

A: uid 和 gid 的 id 都是从 0 开始, 但是 embedding 矩阵 是很大的一个矩阵, 包含了所有 id 的 embedding

Q: 对于特征中  $\text{id} < 0$  的 (譬如  $\text{id} = -1$ ) 是异常数据吗?

A:  $\text{id} < 0$  就是没有这个特征

Q: 人脸属性数据中一个视频里有多个人脸的情况是指视频中出现的所有人, 每个人挑了一个框, 还是同一个人会在不同帧里记录不同的框?

A: 人脸属性数据取的视频中某一帧的所有人脸

Q: 请问 train 文件中 create\_time 字段的编码方式是什么? 是相对哪一天转换的?

A: train 文件中 create\_time 字段的编码方式是秒; 时间是经过脱敏的, 可以根据相对时间看。测试集中 creat\_time 没有在训练集的范围內。

## 1 月 21 日更新

**Q: 分为两个赛道，两赛道数据集是否不同**

**Q1:** track1 提供大规模数据集，track2 提供小规模数据集，track2 的数据是否为 track1 的数据随机抽取出来的？

**A1:** 两个赛道的数据集完全独立，没有任何重合。相应的测试集也是完全独立。

**Q2:** 如果两个 track 数据集不一样，是否可以两个 track 数据集都下载丰富训练集？

**A2:** 两个赛道的数据集完全独立，测试集也相应地完全独立，故使用两个赛道数据集同时进行训练并不能起到丰富训练集的效果。

**Q3:** 是否可以两个 track 都做，看哪个得分高？或者参赛过程中，更换赛道？

**A3:** 参赛队可以视计算资源情况选择任何赛道训练模型并提交结果，因两个赛道的数据集完全独立，两个赛道的榜单也是完全独立的，故各自计算成绩和排名，无需更换赛道。

**Q:** 是否可以使用其他数据？如公共 API 抓取，Web 抓取，或两者兼而有之？

**A:** 竞赛规则中明确指出：“不允许使用主办方提供的数据集之外的任何外部标注数据”，一经发现，做退赛处理。