

# 文件说明/Read me

## 交互数据

示例:

```
uid \t user_city \t item_id \t author_id \t item_city \t channel \t finish \t like \t music_id \t
did \t creat_time \t video_duration
57384      52      43192      142828      0      0      0      0
4513      34178      53085993699      39
```

对应文件:

**final\_track2\_train.txt**

**final\_track1\_train.txt**

## nlp 特征:

获取方式: 视频标题 -> 去除数字, 字母特殊字符, -> jieba 分词 -> 脱敏 成 数字

示例:

```
{"item_id": 4036886, "title_features": {"a": b}}
```

a 是 word id, b 是改 word id 在该句子中出现的次数

```
{"item_id": 4036886, "title_features": {"1": 1, "2": 1, "3": 1, "4": 1, "5": 1, "6": 1}}
```

对应文件:

**track2\_title.txt**

**track1\_title.txt**

## 视觉特征:

示例:

```
{"item_id": 11274473, "video_feature_dim_128": [0, 128]}
```

对应文件:

**track2\_video\_features.txt**

**track1\_video\_features\_part[1-11].txt**

## 人脸特征:

示例:

```
{"item_id": 6603879, "face_attrs": [{"gender": 0, "beauty": 0.53, "relative_position":
[0.4306, 0.3203, 0.3333, 0.2969]}]}
```

性别, 0, 女, 1, 男

相对位置: 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

对应文件:

**track2\_face\_attrs.txt**

**track1\_face\_attrs.txt**

音频特征（2 月中旬提供）

字段说明

字段	字段描述	数据类型	备注
uid	用户 id	int	已脱敏
user_city	用户所在城市	int	已脱敏
item_id	作品 id	int	已脱敏
author_id	作者 id	int	已脱敏
item_city	作品城市	int	已脱敏
channel	观看到该作品的来源	int	已脱敏
finish	是否浏览完作品	bool	
like	是否对作品点赞	bool	
music_id	音乐 id	int	已脱敏
device	设备 id	int	已脱敏
time	track1：作品观看的起始时间 track2：作品发布时间	int	已脱敏
duration time	作品时长	int	单位：秒