

ModelArts 平台使用指导

## 交通流量预测 baseline

文档版本 01  
发布日期 2020-04-17

华为技术有限公司



**版权所有 © 华为技术有限公司 2019。 保留一切权利。**

非经本公司书面许可，任何单位和个人不得擅自摘抄、复制本文档内容的部分或全部，并不得以任何形式传播。

## 商标声明



HUAWEI和其他华为商标均为华为技术有限公司的商标。

本文档提及的其他所有商标或注册商标，由各自的所有人拥有。

## 注意

您购买的产品、服务或特性等应受华为公司商业合同和条款的约束，本文档中描述的全部或部分产品、服务或特性可能不在您的购买或使用范围之内。除非合同另有约定，华为公司对本文档内容不做任何明示或默示的声明或保证。

由于产品版本升级或其他原因，本文档内容会不定期进行更新。除非另有约定，本文档仅作为使用指导，本文档中的所有陈述、信息和建议不构成任何明示或暗示的担保。

## 华为技术有限公司

地址：                    深圳市龙岗区坂田华为总部办公楼                    邮编：518129

网址：                    <http://www.huawei.com>

客户服务邮箱：          [support@huawei.com](mailto:support@huawei.com)

客户服务电话：          4008302118

# 目 录

---

1 文档导读 .....	1
2 准备工作 .....	2
2.1 在 OBS 创建桶及文件夹 .....	2
2.2 在 ModelArts 全局配置中添加访问密钥 .....	3
3 在 ModelArts Notebook 编写训练脚本并且完成训练.....	5
4 将生成的模型导入至模型管理 .....	9
5 将模型部署为批量服务 .....	10
6 提交竞赛结果 .....	12
A 附录 .....	13

# 1 文档导读

ModelArts 是面向 AI 开发者的一站式开发平台，您可以使用 ModelArts 完成交通流量预测赛题数据的分析、训练建模、推理测试以及结果的提交。

本文档包含使用 ModelArts 完成交通流量预测赛题的全流程指导，通过本文档您可以快速了解 ModelArts 平台的使用，有助于您使用 ModelArts 来完成交通流量预测赛题。

您可以根据表 1-1 快速查找您需要了解的内容。

表1-1 文档导读

章节	说明
<a href="#">2 准备工作</a>	创建 OBS 桶以及文件夹、完成 ModelArts 配置。
<a href="#">3 在 ModelArts Notebook 编写训练脚本并且完成训练</a>	使用 ModelArts Notebook 编写模型训练脚本并在 Notebook 中完成训练。
<a href="#">4 将生成的模型导入至模型管理</a>	将训练生成的模型导入至模型管理。
<a href="#">5 将模型部署为批量服务</a>	将模型部署为批量服务，查看预测结果。
<a href="#">6 提交竞赛结果</a>	将模型提交至竞赛指定的账号进行评分。
<a href="#">A 附录</a>	本 baseline 中的训练脚本、推理代码以及批量预测数据。

# 2 准备工作

ModelArts 使用对象存储服务（OBS）存储数据和模型的备份，实现安全、高可靠和低成本的存储需求。

竞赛数据集已上传至华为云对象存储服务（OBS）公共桶中，您首先需要在 OBS 创建一个属于您的空桶及文件夹，然后完成 ModelArts 配置，以完成准备工作。

## 2.1 在 OBS 创建桶及文件夹

首先在 OBS 创建桶及文件夹，用于存放竞赛数据集、模型训练输出的模型文件及部署批量服务时的预测文件，具体操作如下。

1. 登录 OBS 管理控制台 <https://storage.huaweicloud.com/obs/?&region=cn-north-1&locale=zh-cn#/obs/buckets>。
2. 单击页面右上方“创建桶”，在弹出的页面中选择区域、自定义桶名称，可参考下图完成创建桶参数配置。本 baseline 创建的桶名称为 obs-mybucket-bj4。

区域: 华北-北京四

不同区域的资源之间互不互通，请选择靠近您业务的区域。可以降低网络时延，提高访问速度。桶创建成功后不支持变更区域，请谨慎选择。

桶名称: obs-mybucket-bj4

命名规则:

- 全局唯一，不能与已有的任何桶名称重复。
- 长度范围为3到63个字符，支持小写字母、数字、中划线（-）、英文句号（.）。
- 禁止两个英文句号（.）或英文句号（.）和中划线（-）相邻，禁止以英文句号（.）和中划线（-）开头或结尾。
- 禁止使用IP地址。
- 如果名称中包含英文句号（.），访问桶或对象时可能会进行安全证书校验。

存储类别: 标准存储 | 低频访问存储 | 归档存储

适用于有大量热点文件或小文件，且需要频繁访问（平均一个月多次）并快速获取数据的业务场景。

上传对象时，对象默认与桶的存储类别相同，也可以根据适用场景修改。 [了解更多](#)

桶策略: 私有 | 公共读 | 公共读写

桶的拥有者拥有完全控制权，其他用户在未经授权的情况下均无访问权限。

默认加密: 当前仅【华北-北京一、华东-上海二、华东-上海一、非洲-约翰内斯堡】区域支持在创建桶时开启默认加密。

归档数据直读: 开启 | 关闭

通过归档数据直读，您可以直接下载存储类别为归档存储的对象，而无需提前恢复。归档数据直读会收取相应的费用。 [价格详情](#)

多AZ: 开启 | 关闭

多AZ能提高您的数据可用性，同时会采用相对较高的计费标准。 [价格详情](#)

桶的多AZ属性一旦启用，后续无法修改。

**按用量收费**  
创建免费，使用阶段按照用量收费。 [了解计费详情](#)

立即创建

### 注意

- “区域”请选择“华北-北京四”。
- OBS 桶名称全局唯一，不能与已有的任何桶名称重复。

- 单击“立即创建”。
- 在 OBS 桶列表单击桶名称，然后单击左侧导航栏中“对象”，单击页面中“新建文件夹”，自定义文件夹名称后，单击“确定”。
  - 新建一个文件夹存放训练数据集；本 baseline 存放数据集路径为/obs-mybucket-bj4/traffic\_flow\_dataset。
  - 新建一个文件夹存放训练输出文件，然后在该文件夹下新建一个存放模型文件的文件夹；本 baseline 存放模型文件路径为/obs-mybucket-bj4/modelfiles/model。
  - 新建一个文件夹存放批量预测的数据文件；本 baseline 存放批量预测数据的路径为/obs-mybucket-bj4/batchin。
  - 新建一个文件夹存放批量预测的结果文件；本 baseline 存放批量预测结果的路径为/obs-mybucket-bj4/result。



### 注意

- 由于在导入模型时，模型文件必须存放于 model 目录下，且选择模型时要选择 model 目录的上一级，因此必须创建 model 文件夹，且不能直接在 OBS 桶根目录创建 model 文件夹。

## 2.2 在 ModelArts 全局配置中添加访问密钥

使用 ModelArts Notebook、训练作业、模型及服务时可能需要用到对象存储服务，因此在使用 ModelArts 之前，您需要添加访问密钥。

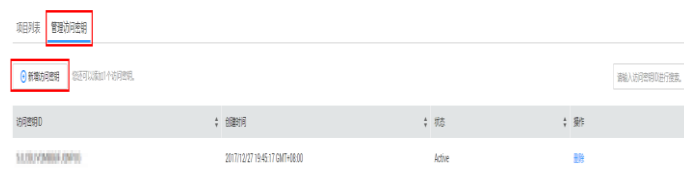
添加访问密钥前，您需要在“我的凭证”页面创建访问密钥，然后在 ModelArts 全局配置中添加访问密钥，具体操作如下。

## 创建访问密钥

1. 登录[华为云](https://console.huaweicloud.com/iam/#/myCredential)，打开“我的凭证”页面（您可直接单击<https://console.huaweicloud.com/iam/#/myCredential>，进入“我的凭证”页面）。



2. 在“我的凭证”页面，单击“管理访问密钥”页签下方的“新增访问密钥”。



3. 在弹出的“新增访问密钥”对话框，输入当前用户的登录密码，通过已验证手机或已验证邮箱进行验证，输入对应的验证码。
4. 单击“确定”，根据浏览器提示，保存密钥文件。密钥文件会直接保存到浏览器默认的下载文件夹中。
5. 打开下载下来的“credentials.csv”文件，即可获取到访问密钥（Access Key Id 和 Secret Access Key）。

## 添加访问密钥

1. 登录 ModelArts 管理控制台。  
在管理控制台上方导航栏，单击“服务列表”，选择“EI 企业智能 > ModelArts”，进入 ModelArts 管理控制台。
2. 在 ModelArts 左侧导航栏选择“全局配置”，单击页面中的“添加访问密钥”，在弹出的对话框中，填写获取的访问密钥。
  - “访问密钥（AK）”输入密钥文件中的 Access Key Id 字段内容，“私有访问密钥（SK）”输入密钥文件中 Secret Access Key 字段内容。
  - 请确保所填写的 AK、SK 为当前账号所获取的。
3. 单击“确认”，完成访问密钥的添加。

# 3 在 ModelArts Notebook 编写训练脚本并且完成训练

ModelArts 集成了 Jupyter Notebook，您可以在 Notebook 中编写训练脚本并在 Notebook 中完成模型的训练；如果您训练的模型需要推理代码，也可以在 Notebook 中编写。

## 在 ModelArts 创建 Notebook

1. 在 ModelArts 左侧导航栏选择“开发环境 > Notebook”，单击页面中“创建”。
1. 在创建 Notebook 页面填写参数，可参考下图填写配置参数。

The screenshot displays the configuration interface for creating a new Notebook in ModelArts. The settings are as follows:

- 计费模式 (Billing Mode):** 按需计费 (Pay-as-you-go)
- 名称 (Name):** notebook-traffic
- 描述 (Description):** (Empty text box)
- 自动停止 (Auto-stop):** Enabled (开关)
- 自动停止时间 (Auto-stop time):** 1小时后 (After 1 hour)
- 工作环境 (Work Environment):** Python3
- 资源池 (Resource Pool):** 公共资源池 (Public Resource Pool)
- 类型 (Type):** CPU
- 规格 (Specification):** 2核 8GiB
- 存储配置 (Storage Configuration):** 云硬盘 (EVS)
- 磁盘规格 (Disk Specification):** 5 GB

Additional information visible in the screenshot includes a character count (0/512) for the description, a warning about auto-stop time, and a note about disk specifications: "磁盘规格默认为5GB, 当磁盘规格为5GB时不收费。超出5GB时, 从Notebook实例创建成功起, 直至删除成功, 超出部分每GB按照规定费用收费。"

2. 单击“下一步”进行规格确认，确认无误后单击“提交”。
3. 单击 Notebook 作业列表操作栏中“打开”，打开 Notebook。
4. 单击“Files”页签右上角“New”，选择对应框架，然后进入代码开发界面。

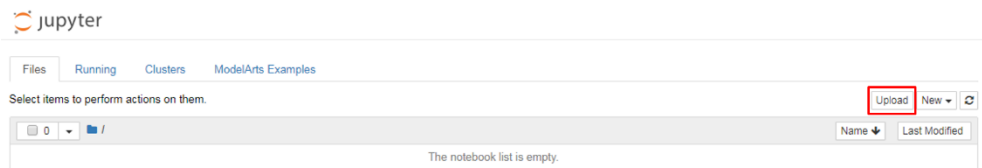


ModelArts 支持 Spark、Scikit-learn、XGBoost 机器学习引擎，您可以基于自身使用习惯来选择；您还可以使用 `pip install` 在 ModelArts Notebook 中安装外部库。

## 编写模型训练脚本及完成训练

本 baseline 已经在 ModelArts Notebook 中编写并调试好模型训练脚本，您可以将代码中数据集路径和模型训练输出路径替换成您创建的 OBS 路径，然后再运行训练脚本，具体操作如下。

1. 将附录中 baseline files.zip 保存至本地，然后解压缩。
2. 打开 Notebook，单击界面中“Upload”将 baseline.ipynb 文件上传至 Notebook。



3. 单击“baseline.ipynb”进入代码开发界面，修改参数“OBS\_DATA\_PATH”和“OBS\_MODEL\_DIR”的值。

本 baseline 选择的框架为 Scikit\_Learn，打开代码文件后选择 XGBoost-Sklernan。

- 将“obs-mybucket-bj4/traffic\_flow\_dataset”改为您的存放竞赛数据的 OBS 桶及文件夹。
- 将“obs-mybucket-bj4/modelfiles/model”改为您的存放模型文件的 OBS 桶以及文件夹。

```
# 获取竞赛数据集：将“obs-mybucket-bj4/myfolder”改成您的OBS桶名及文件夹
import moxing as mox
mox.file.copy_parallel('s3://obs-bdc2020-bj4/traffic_flow_dataset', 's3://obs-mybucket-bj4/traffic_flow_dataset')
print('Copy procedure is completed !')

OBS_DATA_PATH = "s3://obs-mybucket-bj4/traffic_flow_dataset"
LOCAL_DATA_PATH = './dataset/train'
OBS_MODEL_DIR = "s3://obs-mybucket-bj4/modelfiles/model"
OBS_MODEL_PATH = OBS_MODEL_DIR + "/modelfile.m"
OBS_CONFIG_PATH = OBS_MODEL_DIR + "/config.json"
LOCAL_MODEL_PATH = './modelfile.m'
LOCAL_CONFIG_PATH = './config.json'

# read data of one day and one direction
def read_file(path, filename):
    calfile = os.path.join(path, filename)
    original = pd.read_csv(calfile, header=None)
    data = pd.DataFrame(columns=["time", "number"])
    data["time"] = original[0]
    data["number"] = original[3] + original[4]
    return data
```

4. 单击“Run”运行代码。

当代码运行完成后，您可以在 OBS 管理控制台查看训练输出路径下模型文件，其中 config.json 为模型配置文件，modelfile.m 为模型文件。

## 编写推理代码

如果您的模型需要推理代码，您可以在 ModelArts Notebook 中编写，编写完成后将推理代码保存并上传至 OBS 存放模型的 model 文件夹下。

本 baseline 已经编写好模型推理代码，您可以将附录中的推理代码 customize\_service.py 上传至您创建的 OBS 桶中（存放模型的 model 文件夹下）。本 baseline 是上传至/ obs-mybucket-bj4/modelfiles/model 目录下。

对象 | 已删除对象 | 碎片

对象是数据存储的基本单位，在OBS中文件和文件夹都是对象。您可以上传任何类型（文本、

上传对象

新建文件夹

恢复

删除

修改存储类别



名称 ▾

存储类别 ▾

大小 ▾

← 返回上一级



customize\_service.py

标准存储

1.25 KB



config.json

标准存储

481 byte



modelfile.m

标准存储

1.48 MB

如果您自己编写推理代码，具体操作可参考如下步骤。

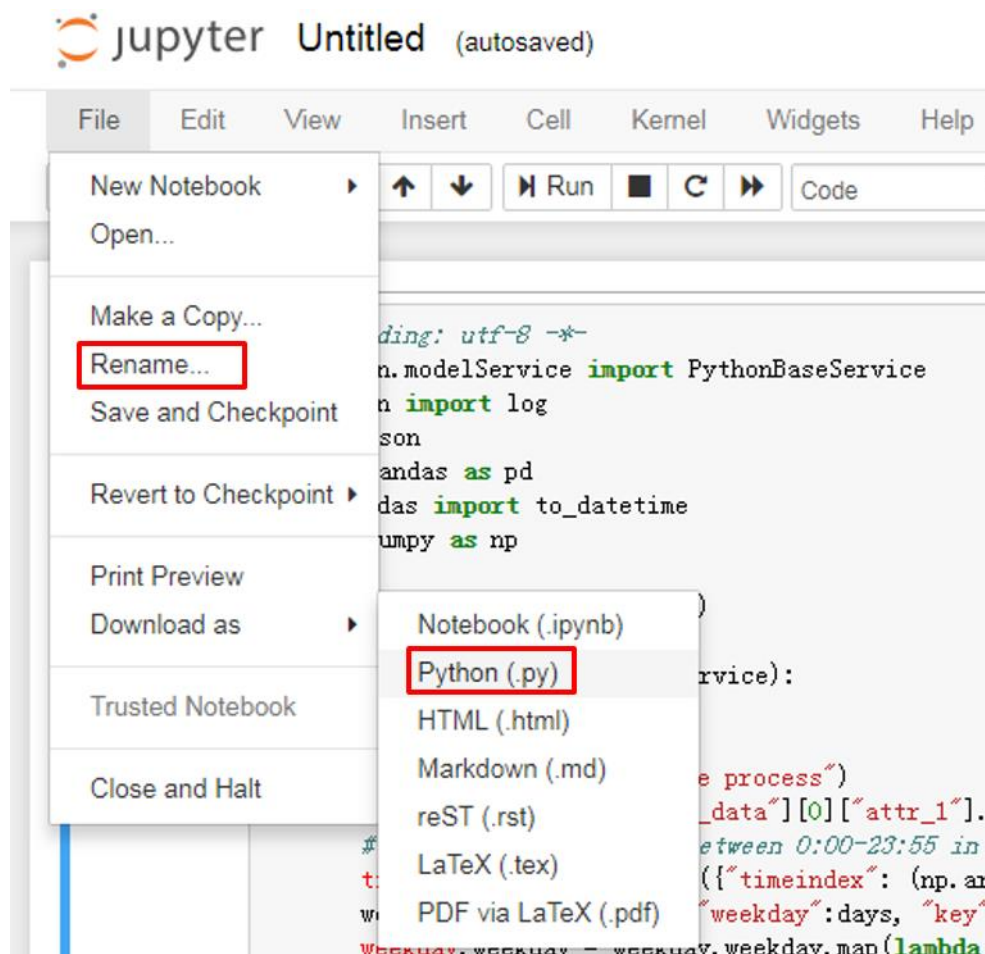
1. 打开 Notebook，单击“Files”页签右上角“New”，选择框架，然后进入代码开发界面，然后您可以进行推理代码的编写。

模型推理代码的编写可参考 [https://support.huaweicloud.com/engineers-modelarts/modelarts\\_23\\_0093.html](https://support.huaweicloud.com/engineers-modelarts/modelarts_23_0093.html)。

2. 编写完成后，单击“File”页签中“Rename”，在弹出的对话框中修改文件名称为“customize\_service”。

注意：推理代码名称固定为 customize\_service，且为.py 类型文件。

3. 单击“File”页签中“Download as > Python(.py)”，将文件下载至本地。

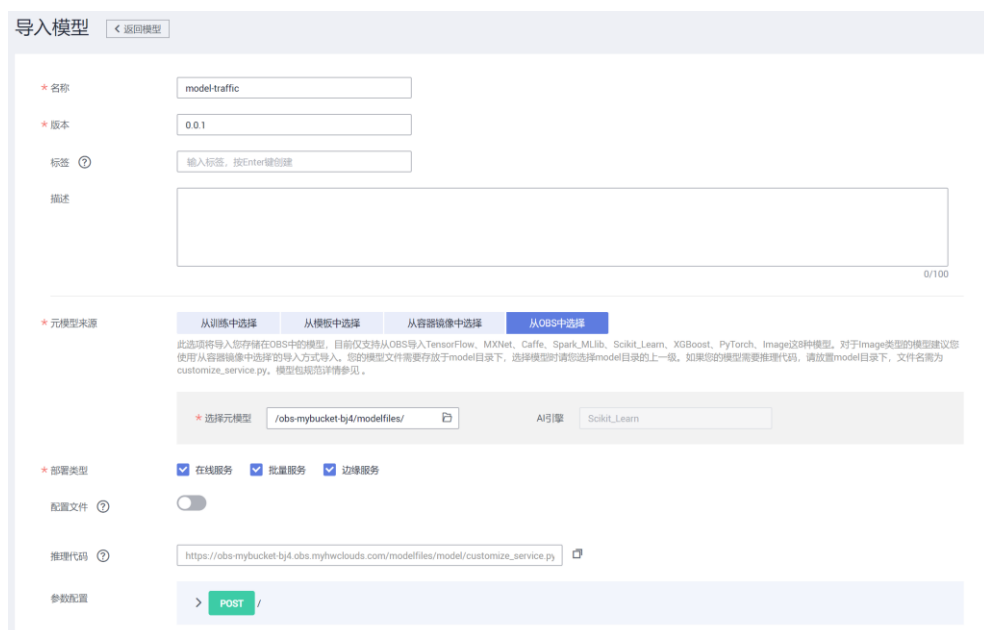


4. 将下载的 customize\_service.py 文件上传至 OBS 存放模型的 model 文件夹下。OBS 上传文件操作可参考 [https://support.huaweicloud.com/usermanual-obs/zh-cn\\_topic\\_0045829661.html](https://support.huaweicloud.com/usermanual-obs/zh-cn_topic_0045829661.html)。

# 4 将生成的模型导入至模型管理

模型训练完成且编写好模型推理代码后，您可以将生成的模型导入至模型管理。导入模型具体操作如下。

1. 在 ModelArts 左侧导航栏选择“模型管理”，单击页面中“导入”。
2. 在导入模型页面填写名称，选择元模型来源；其中元模型来源如从 OBS 中选择，请选择 model 文件夹上一级目录；本 baseline model 文件夹上一级目录为/obs-mybucket-bj4/modelfiles/。



3. 单击“立即创建”。当模型版本状态为“正常”后，即导入模型成功，然后进行下一步将模型部署为批量服务的操作。

说明：也可以部署为批量服务。

# 5 将模型部署为批量服务

---

导入模型完成后，接下来您可以将模型部署为批量服务。

部署之前您需要将批量预测的.csv 数据文件上传至 OBS 桶中，本 baseline 是将批量预测文件 batchin.csv 上传至/obs-mybucket-bj4/batchin 目录下，其中 batchin.csv 请从附录文件中获取。



## 说明

- 批量预测数据文件类型为.csv。
- 本 baseline 是预测 2019-2-11 和 2019-2-14 两天的交通流量，如需要预测其他日期的交通流量，可直接修改 batchin.csv 文件中的时间。

部署为批量服务具体步骤如下。

1. 在 ModelArts 左侧导航栏中选择“部署上线 > 批量服务”，然后单击页面中的“部署”。
2. 在部署页面填写参数，其中“输入数据目录位置”选择存放批量预测数据文件的 OBS 路径。

部署

< 返回批量服务

\* 名称

service-traffic

描述

0/100

\* 模型来源

我的模型

市场订阅

\* 选择模型及版本

model-traffic-flow

0.0.1

\* 输入数据目录位置 ?

/obs-mybucket-bj4/batchin/

\* 请求路径 ?

/

\* 映射关系 ?

```
5      "type": "array",
6      "items": [
7        {
8          "type": "string",
9          "index": 0
10       }
11     ]
12   }
13 }
14 }
```

\* 输出数据目录位置 ?

/obs-mybucket-bj4/result/

\* 计算节点规格

CPU: 2 核 8 GiB

适合场景：CPU标准规格，满足大多数模型的运行和预测

选择“输出数据目录位置”时，可在弹出的对话框中新建文件夹来存放批量预测结果。本 **baseline** 是将预测结果存放在 **result** 文件夹里。

- 单击“立即创建”。
- 当批量服务状态变为“运行完成”后，您可以在 OBS 管理控制台查看输出的预测结果，即/obs-mybucket-bj4/result 中生成的 batchin.csv\_result.txt 文件，可下载至本地然后查看预测结果。

说明：如果您部署的是在线服务，部署完成后，在预测页签输入{"req\_data":["2019-2-11,2019-2-14"]}，在页面右侧即可查看预测结果。

# 6 提交竞赛结果

模型部署测试无误后，您可以将模型提交评审，具体操作如下。

1. 在 ModelArts 左侧导航栏中选择“模型管理>模型”，单击模型名称左侧“√”，然后单击页面右侧操作栏中的“发布>参赛发布”。



2. 在弹出的“参赛模型提交”对话框中，选择比赛项目、比赛阶段，然后单击确定。

点击确定后，即成功提交模型判分。在如下界面中可点击“现在加入”，也可以点击“以后再说”或直接点击右上角关掉该对话框。

## 参赛模型提交



您的参赛模型提交申请已成功受理

小M推荐您加入AI市场，与其他参赛小伙伴，一起分享算法经验。甚至还可以分享和交易自己开发的模型。这是一个有趣的社区，将会包含数据集、案例集、模型算法等丰富的资源。现在加入，即可获得参赛模型提交结果，判分结果的邮件实时通知哦！

我是ModelHub,你的小M，我为AI市场代言

现在加入

以后再说

说明：模型提交判分后，需等待一定时间（判分系统进行判分需一定时间，运行时长与选手提交的模型有关），判分系统完成判分后，可在竞赛平台“提交作品”中查看得分，其中“提交作品”页面需报名比赛后才会显示。

# A 附录

---

如下 baseline files.zip 文件包含了本 baseline 中以下文件：

- baseline.ipynb: 模型训练脚本
- customize\_service.py: 模型推理代码
- batchin.csv: 批量预测数据文件



baseline files.zip