



深圳市数据开放创新大赛 - ModelArts 平台使用指导

交通流量预测 baseline

文档版本 01
发布日期 2019-05-14

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

客户服务电话： 4008302118

目 录

1 文档导读.....1

2 注册华为云账号.....2

3 获取竞赛数据集.....3

3.1 在 OBS 创建桶及文件夹.....3

3.2 在 ModelArts 全局配置中添加访问密钥5

3.3 在 ModelArts 创建 Notebook 并通过代码拷贝竞赛数据集.....6

4 在 ModelArts Notebook 编写训练脚本并且完成训练8

5 将生成的模型导入至模型管理12

6 将模型部署为批量服务13

7 提交竞赛结果.....16

A 附录18

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
2 注册华为云账号	使用 ModelArts 前需注册华为云账号并进行实名认证；如已有华为云账号可跳过此步骤。
3 获取竞赛数据集	获取交流流量预测赛题数据集。
4 在 ModelArts Notebook 编写训练脚本并且完成训练	使用 ModelArts Notebook 编写模型训练脚本并在 Notebook 中完成训练。
5 将生成的模型导入至模型管理	将训练生成的模型导入至模型管理。
6 将模型部署为批量服务	将模型部署为批量服务，查看预测结果。
7 提交竞赛结果	将模型提交至竞赛指定的账号进行评分。
A 附录	本 baseline 中的训练脚本、推理代码以及批量预测数据。

2 注册华为云账号

使用华为云服务之前您需要注册华为云帐号，通过华为云帐号，您可以使用所有华为云服务。注册华为云账号具体操作如下。

1. 进入[华为云](#)首页，单击页面右上角的“注册”。
2. 设置手机号、短信验证码、账号名、密码并勾选“我已阅读并同意《华为云用户协议》和《隐私政策声明》”，单击“同意协议并注册”。
3. 账号注册成功后，登录华为云，单击界面右上角“用户名>实名认证”，根据界面提示信息完成实名认证。

3 获取竞赛数据集

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；然后在 ModelArts 开发环境中创建 Notebook；最后在 Notebook 中执行两行简单的代码，将竞赛数据集拷贝至您创建的 OBS 桶中。

3.1 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/?®ion=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-baseline。

创建桶

返回桶列表

区域

华北-北京一

不同区域的资源之间内网互不连通，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-baseline

命名规则：

- 需全局唯一，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。

多AZ

当前仅【华东-上海二、华北-北京四、西南-贵阳一】区域支持配置多AZ。

文件语义

开启

关闭

标签

如果您需要使用同一标签标识多种云资源，即所有服务均可在标签输入框下拉选择同一标签，建议在TMS中创建预定义标签。
 [查看预定义标签](#)

标签键

标签值

按用量收费

创建免费，使用阶段按照用量收费。
 [了解计费逻辑](#)

立即创建

注意

- “区域”请选择“华北-北京一”。
- 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”。新建文件夹详细操作可参见https://support.huaweicloud.com/usermanual-obs/zh-cn_topic_0045829103.html。
 - 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-baseline/data。
 - 新建一个文件夹存放代码文件；本 baseline 存放代码文件路径为 obs-baseline/codes。
 - 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-baseline/output/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-baseline/batchin。

文档版本 01 (2019-05-14)
 版权所有 © 华为技术有限公司
 4

桶列表 > obs-baseline



桶列表 > obs-baseline > output



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。

3.2 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。

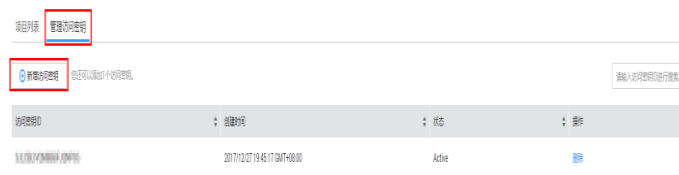
添加访问密钥前，您需要在“我的凭证”页面创建访问密钥，然后在 ModelArts 全局配置中添加访问密钥，具体操作如下。

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。

3.3 在 ModelArts 创建 Notebook 并通过代码拷贝竞赛数据集

本节介绍如何在 ModelArts 开发环境创建 Notebook，并在 Notebook 中通过两行代码将竞赛数据集拷贝至您的 OBS 桶中，具体操作如下。

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

注意

- ModelArts 支持 Spark、Scikit-learn、XGBoost 机器学习引擎，“AI 引擎”可以基于您自身使用习惯来选择。本 baseline 选择的 AI 引擎为 Scikit-learn。
- “存储配置”建议选择“OBS”，并选择您创建的存放代码文件的 OBS 路径，这样您在 Notebook 编写的代码文件会自动保存在该路径下。本 baseline 存放代码的 OBS 路径为 /obs-baseline/codes。

* 计费模式

* 名称

描述
0/256

* AI引擎

* 资源池

* 类型

* 规格

存储配置

* 存储位置 ?

- 单击“下一步”进行规格确认，确认无误后单击“立即创建”。
- 单击 Notebook 作业列表操作栏中“打开”，打开 Notebook。
- 单击“Files”页签右上角“New”，选择“Python 2”进入代码开发界面。

在代码输入栏中输入如下代码，即可从 OBS 公共桶中将竞赛数据拷贝至您创建的 OBS 桶中。

```
import moxing as mox
mox.file.copy_parallel('s3://modelarts-cn-north1-learning-course/dataset/HC-SZ ODC
2019/tfp2019/unarchiver/traffic_flow_prediction_data', 's3://obs-baseline/data')
print('Copy procedure is completed !')
```

说明

- 将“obs-baseline/data”替换为您创建的存放竞赛数据集的 OBS 路径。
- MoXing 是华为云 ModelArts 团队自研的分布式训练加速框架，并提供了一套文件对象 API，可以用来读写 ModelArts Notebook 容器本地文件及 OBS 文件。

4 在 ModelArts Notebook 编写训练脚本并且完成训练

获取竞赛数据之后，您可以在 ModelArts 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。

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。



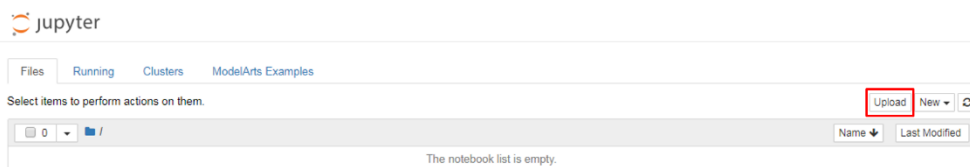
- 如何在 ModelArts Notebook 中安装外部库，可参见《ModelArts 常见 FAQ-V1.0》。

编写模型训练脚本及完成训练

打开 Notebook 后，单击“Files”页签右上角“New”，选择“Python 2”进入代码开发界面，然后您可以进行模型训练脚本的编写和运行。

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。
 - OBS_DATA_PATH 填写您存放训练数据集的 OBS 路径，本 baseline 路径为“s3://obs-baseline/data”。
 - OBS_MODEL_DIR 填写您存放训练输出模型文件的 OBS 路径，本 baseline 路径为“s3://obs-baseline/output/model”。

```
import mxing as mx
mx.file.shift('os', 'mx')
import os
import re
import json
import pandas as pd
from pandas import to_datetime
from sklearn.ensemble import GradientBoostingRegressor
from sklearn.metrics import mean_squared_error
from sklearn.model_selection import train_test_split
from sklearn.externals import joblib
from collections import OrderedDict

OBS_DATA_PATH = "s3://obs-baseline/data"
LOCAL_DATA_PATH = '/home/jovyan/ai-competition/data/train'
OBS_MODEL_DIR = "s3://obs-baseline/output/model"
OBS_MODEL_PATH = OBS_MODEL_DIR + "/modelfile.m"
OBS_CONFIG_PATH = OBS_MODEL_DIR + "/config.json"
LOCAL_MODEL_PATH = '/home/jovyan/ai-competition/modelfile.m'
LOCAL_CONFIG_PATH = '/home/jovyan/ai-competition/config.json'
```

4. 单击“Run”运行代码。

当输出“Model training has been completed!”表示运行成功，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。

桶列表 > obs-baseline > output > model

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

上传对象	新建文件夹	恢复	删除
<input type="checkbox"/> 名称 ▾			
← 返回上一级			
<input type="checkbox"/> config.json			
<input type="checkbox"/> modelfile.m			

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-baseline/output/model` 目录下。

桶列表 > obs-baseline > output > model



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“Python 2”进入代码开发界面，然后您可以进行推理代码的编写。

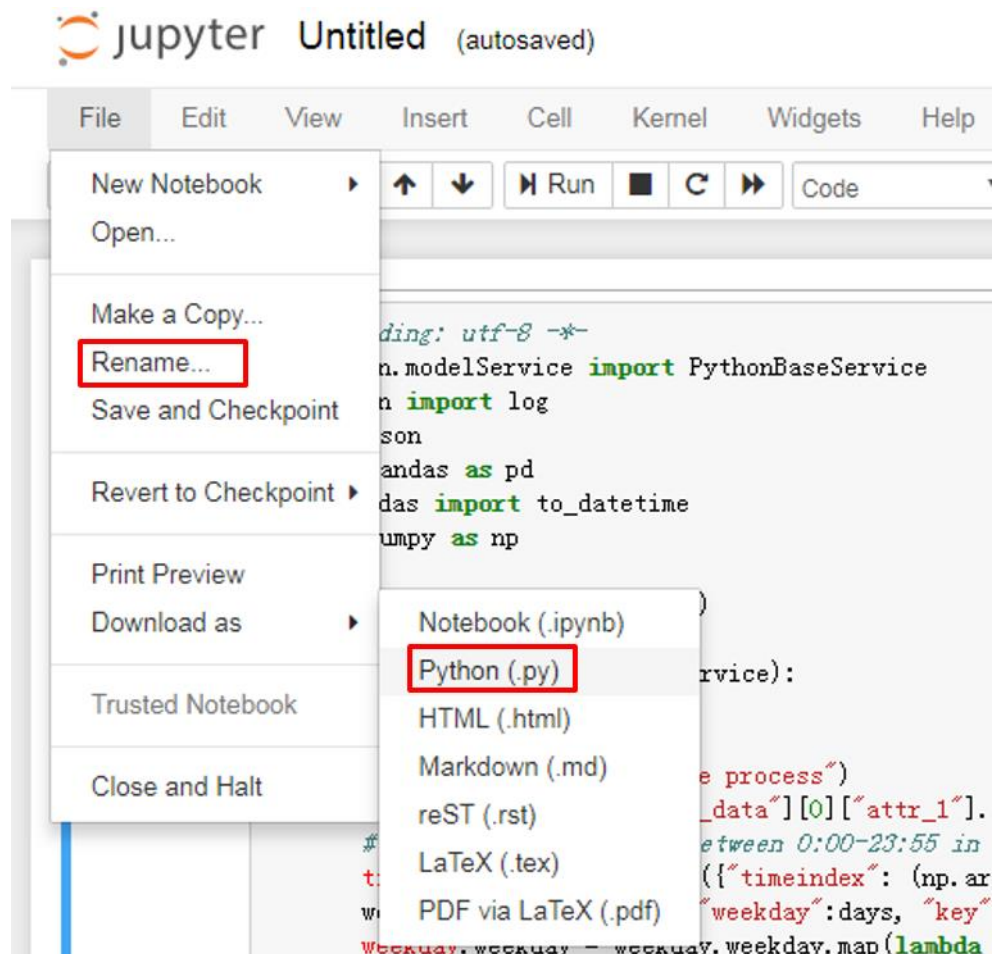
模型推理代码的编写可参考 https://support.huaweicloud.com/usermanual-modelarts/modelarts_02_0061.html。

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。

5 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-baseline/output。

* 名称

model-baseline

* 版本

0.0.1

描述

0/100

* 元模型来源

从训练中选择

从OBS中选择

从容器镜像中选择

此选项将导入您存储在OBS中的模型，目前仅支持从OBS导入TensorFlow、MXNet、Spark_MLlib、Scikit_Learn、XGBoost、MindSpore、Image、PyTorch这8种模型。对于Image类型的模型建议您使用从容器镜像中选择的导入方式导入。您的模型文件需要存放于model目录下，选择模型时请选择model目录的上一级。如果您的模型需要推理代码，请放置model目录下，文件名需为customize_service.py。模型包规范详情参见[模型包规范](#)。

* 选择元模型

/obs-baseline/output/

AI引擎

Scikit_Learn

配置文件

推理代码

https://obs-baseline.obs.myhwclouds.com/output/model/customize_servic

- 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。

6 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-baseline/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 2019-2-4 和 2019-2-7 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，本 baseline 是存放在/obs-baseline/batchin 路径下。



7 提交竞赛结果

模型部署测试后，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”，然后单击页面右侧操作栏中的“市场发布”。

model-baseline

创建新版本

溯源图

请输入版本查询

Q

↺

版本	状态	模型大小	创建时间	描述	操作
0.0.1	<div>正常</div>	1.12 MB	2019/05/06 20:07:50 GMT+08:00	-/✎	<div>部署</div> <div>市场发布</div> <div>删除</div>

2. 在发布模型页面填写参数，其中“发布到”选择“个人”，填写评审账号 ID：**d8126a20db13499c82e060007d1e8348**，然后单击“添加”。

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