



垃圾分类挑战赛 - ModelArts 平台使用指导

Baseline

文档版本 01
发布日期 2019-07-30

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.....3

3 准备数据.....4

4 在 ModelArts 全局配置中添加访问密钥.....8

5 在 Notebook 中编写训练脚本9

6 创建训练作业.....14

7 将生成的模型导入至模型管理16

8 将模型部署为在线服务17

9 提交竞赛结果.....1

10 释放资源.....3

A 附录4

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 完成垃圾分类挑战赛数据处理、训练建模、推理测试以及结果的提交。

本文档包含使用 ModelArts 完成垃圾分类挑战赛的全流程指导，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 来完成垃圾分类挑战赛题目。

您可以根据如下表格快速查找您需要了解的内容。

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 准备数据	下载竞赛数据集，并将数据集上传至 OBS 您的 OBS 桶中。
4 在 ModelArts 全局配置中添加访问密钥	使用 ModelArts 前，添加访问密钥。
5 在 Notebook 中编写训练脚本（可选）	在 ModelArts Notebook 中编写模型的训练脚本。 可选是指：运行 baseline 代码时可跳过此步骤。
6 创建训练作业	基于 ModelArts 提供的 CPU、GPU 模型训练环境，完成模型的训练。
7 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。
8 将模型部署为批量服务	将模型部署为在线服务，并查看预测结果是否正确。
9 提交竞赛结果	将模型提交至竞赛指定的账号，并在竞赛平台上提交作品进行评分。
10 释放资源	为避免资源浪费或账号欠费，使用完计算资源后需及时释放资源。

章节	说明
A 附录	本 baseline 中的训练脚本、推理代码。

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 准备数据

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。因此，在使用 ModelArts 开发模型前，您需要将竞赛数据集上传至您的 OBS 桶中。

本节介绍下载竞赛数据集至本地并解压后，然后通过 OBS Browser 创建桶及文件夹，将竞赛数据集上传至您的 OBS 桶中。具体操作如下。

步骤一：下载竞赛数据集至本地并解压缩

通过如下链接或竞赛平台赛题说明页面下载数据集 `garbage_classify.zip`：

[https://modelarts-competitions.obs.cn-north-](https://modelarts-competitions.obs.cn-north-1.myhuaweicloud.com/garbage_classify/dataset/garbage_classify.zip)

[1.myhuaweicloud.com/garbage_classify/dataset/garbage_classify.zip](https://modelarts-competitions.obs.cn-north-1.myhuaweicloud.com/garbage_classify/dataset/garbage_classify.zip)。

将数据集 `garbage_classify.zip` 下载至本地后解压缩，`garbage_classify` 包含如下文件。

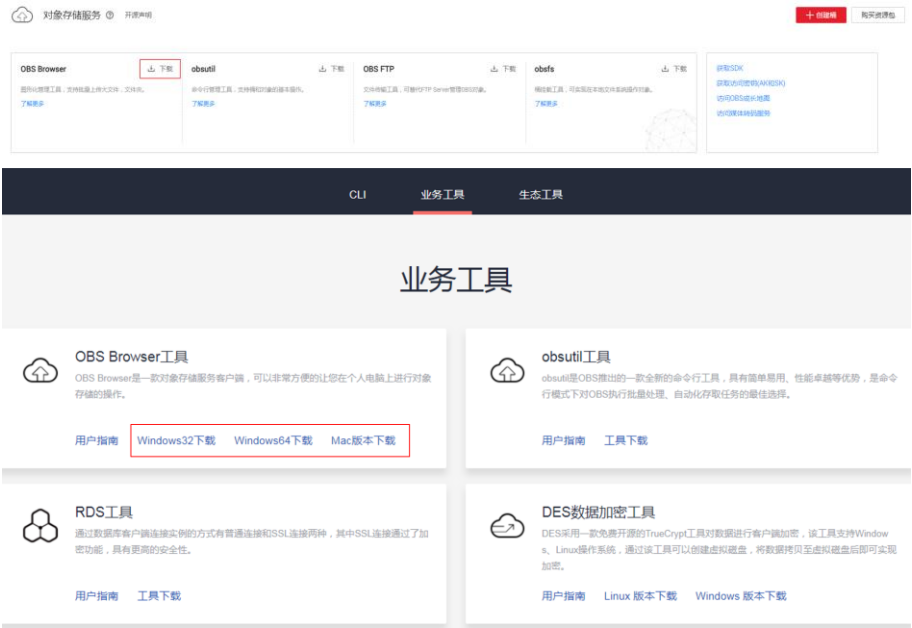
名称	说明
<code>train_data</code>	训练集目录，包含垃圾图片和对应的标签文件（.txt）。
<code>garbage_classify_rule.json</code>	垃圾分类规则字典，key 值是 id，value 是“垃圾种类/具体物品名”。 例如训练数据标签文件 <code>img1.txt</code> 的内容是“ <code>img_1.jpg, 0</code> ”，表示 <code>img_1.jpg</code> 这张图中的垃圾是“其他垃圾/一次性快餐盒”。

说明：您可以根据实际需要将数据集 `train_data` 划分部分数据作为验证集和测试集。

步骤二：下载 OBS Browser 并上传数据集

首先在 OBS 管理控制台下载 OBS Browser，然后在 OBS Browser 创建桶以及文件夹，并将加压缩的数据集上传至 OBS 桶。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. 单击页面中 OBS Browser 下载按钮，并在弹出的页面中选择对应版本。



- 3. 下载完成后解压缩，单击 `obs.exe` 运行 OBS Browser。
- 4. 获取访问密钥（AK 和 SK），具体操作可参见 https://support.huaweicloud.com/modelarts_faq/modelarts_05_0004.html。
- 5. 在 OBS Browser 添加新账号页面自定义账号名，添加获取的访问密钥，然后单击确定。
- 6. 账号添加完成后，单击“添加桶”，区域选择“华北-北京一”，自定义桶名称。
注意：OBS 桶名全局唯一，如提示桶名已存在，则需要更换其他名称。

添加桶
✕

方式

☒ 创建新桶
 ☐ 添加外部桶

区域

华北-北京一
▼
?

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。

桶策略

☒ 私有
 ☐ 公共读
 ☐ 公共读写

桶的拥有者拥有完全控制权限，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-贵阳二】区域支持配置多AZ。

桶名

myobsbucket
?

确定

取消

7. 单击桶名称，然后单击“上传”，在弹出的页面中选择上传文件夹，然后选择解压的数据集（garbage_classify），单击“确定”。

您可以在“任务管理”中查看上传进度。

上传对象
✕

您可以同时上传多个文件和文件夹，一次最多可以上传500个文件和文件夹。

+ 选择文件

+ 选择文件夹

对象数 0

大小 0

对象名称	本地路径	大小	操作

存储类别

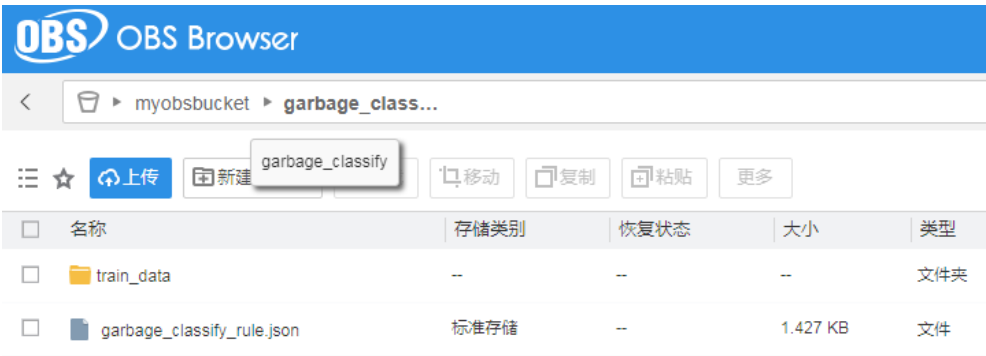
☒ 标准存储
 ☐ 低频访问存储
 ☐ 归档存储

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。

增量上传

☐ 开启
 ☒ 关闭

关闭增量上传后，上传对象时会覆盖桶中同名对象。



4 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，您需要在 ModelArts 全局配置中添加访问密钥。

在步骤 3 准备数据中已创建过访问密钥，可直接将该密钥添加至 ModelArts。具体操作如下。

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。

5 在 Notebook 中编写训练脚本（可选）

标题中可选是指：运行 **baseline** 代码时可不执行本章节内容，可直接 [6 创建训练作业](#) 执行操作。

获取竞赛数据之后，您可以在 ModelArts Notebook 中编写模型训练脚本；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 TensorFlow、MXNet、PyTorch 等深度学习引擎，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。

说明

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

创建 Notebook

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

注意

- 为方面在 ModelArts Notebook 中调试代码，“存储配置”建议选择“EVS”。

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-ljfl
- 描述 (Description):** (Empty text box)
- 工作环境 (Working Environment):** Python2, Python3 (Python3 is selected)
- 资源池 (Resource Pool):** 公共资源池 (Public Resource Pool), 专属资源池 (Dedicated Resource Pool) (Public Resource Pool is selected)
- 类型 (Type):** CPU, GPU (CPU is selected)
- 规格 (Specification):** 2核8GiB, 8核32GiB (2核8GiB is selected)
- 存储配置 (Storage Configuration):** EVS, OBS (EVS is selected)
- 磁盘规格 (Disk Specification):** - 5 + GB. Minimum value: 5; Maximum value: 4096. A note at the bottom states: "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。"

3. 单击“下一步”进行规格确认，确认无误后单击“提交”。

Notebook 创建完成后，可单击 Notebook 作业列表操作栏中“打开”，然后进入 Jupyter Notebook。

编写模型训练脚本

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

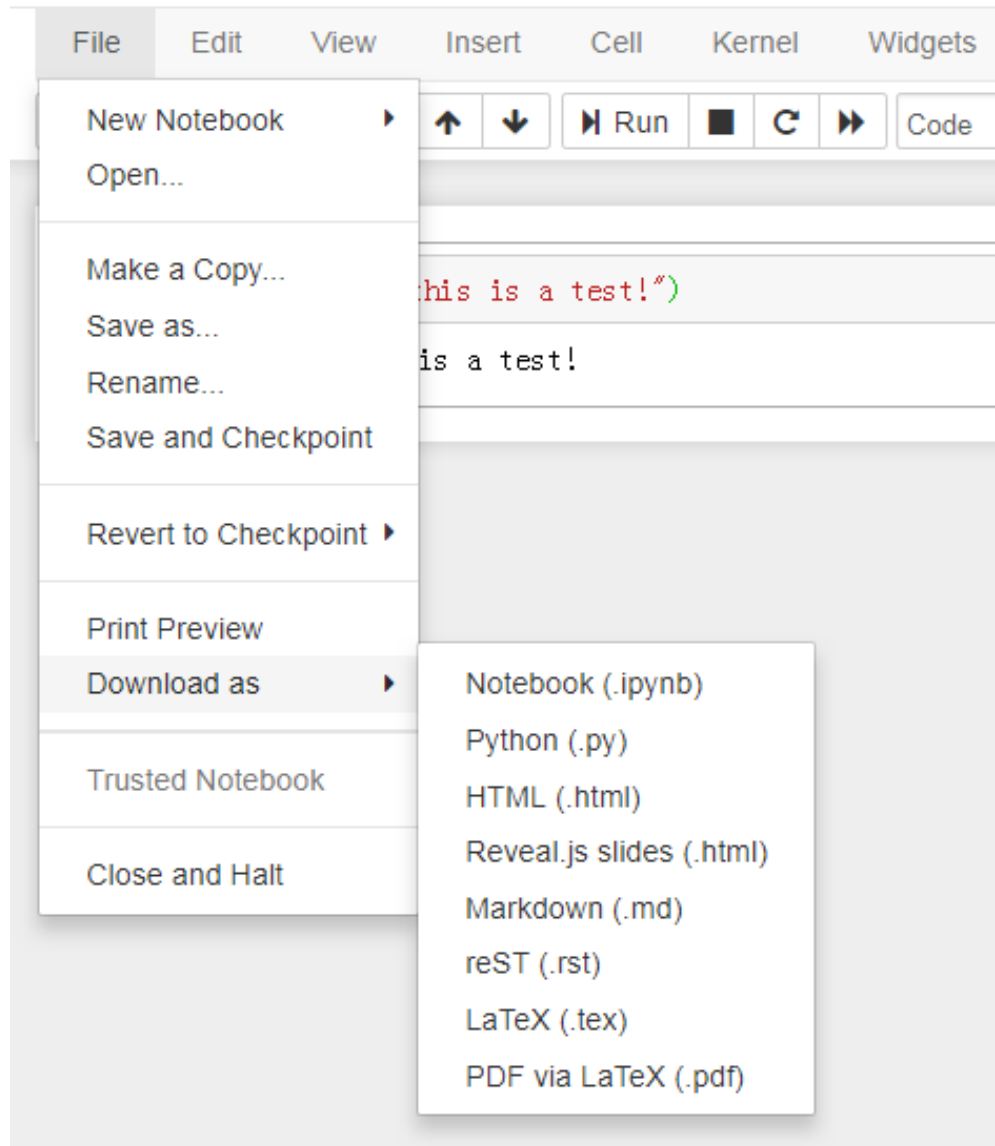
说明：ModelArts 支持 TensorFlow、MXNet、PyTorch 等深度学习引擎，您可以基于自身使用习惯来选择。

本 baseline 已经在 ModelArts Notebook 中编写并调试好模型训练脚本，即附录中的 src.zip，您可以将该压缩包解压后上传至 OBS 桶，然后[创建训练作业](#)。



如果您自己编写训练脚本，具体操作可参考如下步骤。

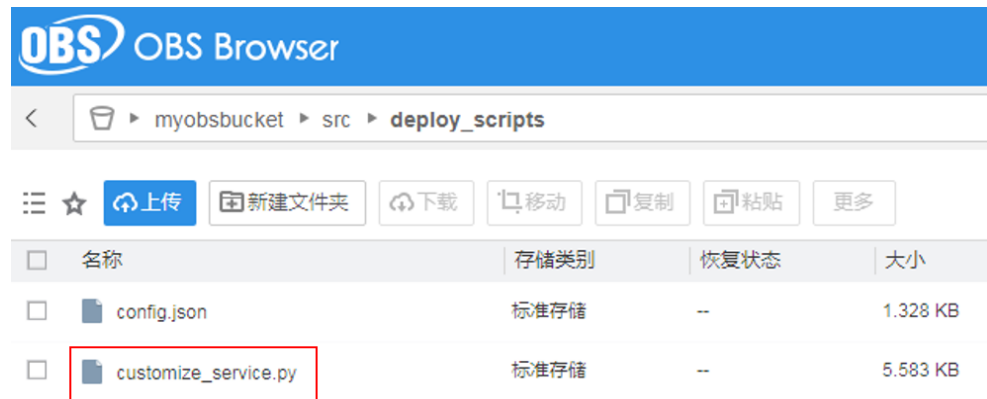
1. 打开 Notebook，单击“Files”页签右上角“New”，选择引擎后进入代码开发界面，然后您可以进行训练脚本的编写。
2. 编写和调试完训练脚本后，需要将当前 ipynb 保存为 Python 文件，才能用于 ModelArts 训练作业。
选择“File > Download as > Python(.py)”，将调试好的代码保存至本地，然后上传至您的 obs 桶，用于创建训练作业。



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 附录 src.zip 包含已编写好的推理代码，如下图所示。



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 3”进入代码开发界面，然后您可以进行推理代码的编写。

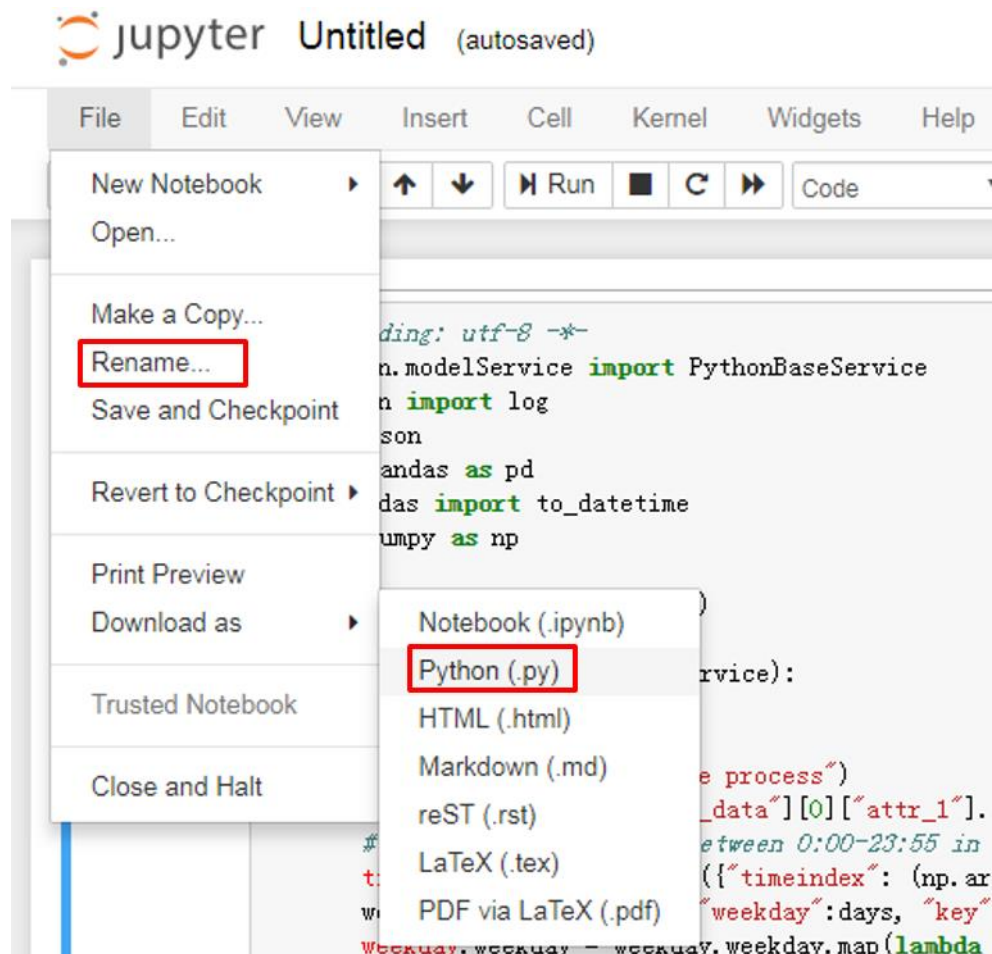
模型推理代码的编写可参考 https://support.huaweicloud.com/engineers-modelarts/modelarts_23_0093.html。

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。

6 创建训练作业

准备好模型训练脚本后，您可以在 ModelArts 训练作业模块创建训练作业进行模型的训练，具体操作如下。

1. 在 ModelArts 左侧导航栏中选择“训练作业”，然后单击页面右侧“训练作业”页签中的“创建”，进入创建训练作业页面。
2. 在创建训练作业页面填写参数，参数填写说明如下。
 - 数据来源：
选择“数据存储位置”，注意 OBS 路径要选择到 `train_data`。
 - 算法来源：
选择“常用框架”，本 baseline“AI 引擎”使用的是 TF-1.13.1-python3.6，“代码目录”选择到 `src`，“启动文件”选择 `run.py`。
 - 运行参数：
`num_classes = 40`，数据的类别。
`deploy_script_path = s3://myobsbucket/src/deploy_scripts`，推理代码&配置文件的路径，注意将桶名称“myobsbucket”替换成您的 OBS 桶。
 - 训练输出位置：
OBS 路径选择到 `model` 文件夹上一层目录，本 baseline 为 `/myobsbucket/output`。
 - 作业日志路径：
您如果配置了“作业日志路径”，可以在训练作业详情“日志”页签下下载日志。

★数据来源 ?

数据集

数据存储位置

★数据存储位置

/myobsbucket/garbage_classify/train_data/

选择

🗑

★算法来源

预置算法

常用框架

自定义

[MoXing手册](#)

选择常用引擎创建训练作业。

★AI引擎

TensorFlow

TF-1.13.1-python3.6

★代码目录 ?

/myobsbucket/src/

选择

★启动文件 ?

/myobsbucket/src/run.py

选择

运行参数 ?

num_classes

=

40

🗑

deploy_script_path

=

s3://myobsbucket/src/deploy_script

🗑

+

增加运行参数

★训练输出位置 ?

/myobsbucket/output/

选择

请尽量选择空目录来作为训练输出路径。

作业日志路径 ?

选择

清除

日志默认保存在服务，会不定期清除，请选择相应路径用来上传日志。

3. 参数填完成后，单击“下一步”，确认参数无误后，单击“提交”。

训练作业运行完成后，在训练输出位置会自动生成 model 文件夹，且 model 文件夹下包含如下文件。

OBS Browser

< 🗑 myobsbucket ▶ output ▶ model

☰

☆

📶 上传

📁 新建文件夹

📶 下载

📁 移动

📄 复制

📄 粘贴

更多

<input type="checkbox"/>	名称	存储类别	恢复状态	大小
<input type="checkbox"/>	📁 variables	--	--	--
<input type="checkbox"/>	📄 config.json	标准存储	--	1.328 KB
<input type="checkbox"/>	📄 customize_service.py	标准存储	--	5.583 KB
<input type="checkbox"/>	📄 saved_model.pb	标准存储	--	1.851 MB

7 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”。

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 /myobsbucket/output。

注意：选择好元模型路径后，“AI 引擎”会自动填充。如未能自动填充，请检查元模型路径是否是 model 上一级目录，或者 model 目录下是否包含模型配置文件 config.json。

* 名称

* 版本

标签

描述

0/100

* 元模型来源

从训练中选择

从模板中选择

从容器镜像中选择

从OBS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、Caffe、Spark_MLib、Scikit_Learn、XGBoost、MindSpore、Image、PyTorch这9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，文件名称为customize_service.py。模型包规范详情参见[模型包规范](#)。

* 选择元模型

AI引擎

配置文件

推理代码

2. 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[将模型部署为在线服务](#)的操作。

model-ijfl

创建新版本

溯源图

请输入版本查询

版本	状态	模型大小	模型来源	创建时间	描述	操作
0.0.1	正常	138.00 MB	自定义算法	2019/08/02 09:15:49 ...	--	部署 市场发布 删除

8 将模型部署为在线服务

导入模型完成后，接下来您可以将模型部署为在线服务。

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 > 在线服务”，然后单击页面中的“部署”。
2. 在部署页面填写参数，其中在“模型列表”选择要导入的模型及版本。

3. 单击“下一步”，参数确认无误后，单击“提交”。
创建成功后，您可以在在线服务列表中查看部署进度，当状态变为“运行中”后，单击服务名称，进入详情页面，如下图所示。

在线服务 > service-ljfl

名称

service-ljfl

服务ID

8b7b8118-d49d-4362-b1b5-7b3e81d494a6

状态

运行中

来源

我的部署

调用失败次数/总次数

0 / 0

网络配置

未设置

描述

--

调用指南

预测

配置更新记录

监控信息

日志

共享

溯源图

API接口地址 <https://dedc1ddd2def4d34a8c47704a05b11e5.ap>

模型：

model-ljfl 0.0.1

参数配置

POST /

输入参数

名称	类型
input_img	file

输出参数

名称	类型
result	string

在“预测”页签添加图片进行测试，如下图所示。

名称

service-ljfl

服务ID

8b7b8118-d49d-4362-b1b5-7b3e81d494a6

状态

运行中

来源

我的部署

调用失败次数/总次数

0 / 0

网络配置

未设置

描述

--

调用指南

预测

配置更新记录

监控信息

日志

共享

溯源图

请求路径：

/

选择预测文件

上传

重新预测

预测图片预览

预测结果显示

预测成功

```
1 {  
2   "result": "其他垃圾/一次性快餐盒"  
3 }
```



9 提交竞赛结果

模型部署测试后，您可以将模型提交评审，首先需要通过 ModelArts 将模型发布给大赛评审账号，然后在竞赛平台上提交作品。

步骤一：在 ModelArts 平台上提交模型

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-ljfl

创建新版本

溯源图

请输入版本查询

版本	状态	模型大小	模型来源	创建时间	描述	操作
0.0.1	正常	138.00 MB	自定义算法	2019/08/02 09:15:49 ...		部署 市场发布 删除

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”。

发布者

?

team-name

名称

team-name-model1.0

描述

?

-

模型画像

?

行业

数据

场景

主题

模型

框架引擎

封面图

?

使用上一次的封面图

OBS

☒ 本地上传

samplephoto.png

选择

发布到

?

AI市场

☒ 个人

如何查看账号id

d8126a20db13499c82e060007d1e8348

添加

已发布给1个账号

确定

取消

3. 单击“确定”，完成模型的提交。

步骤二：在竞赛平台上提交模型

在大赛平台上点击提交作品-上传作品，选择已发布给大赛指定账号的模型，其中“提交作品”页面需报名比赛后才会显示。

赛事介绍	上传作品			
	*为了竞赛公平性，平台统一使用modelarts实现开发			
赛题说明				
提交作品	作品名称	阶段	提交时间	得分

注意

- 要先在 ModelArts 上提交模型后，然后到竞赛平台上“提交作品”，且顺序不能反。

10 释放资源

使用完 ModelArts 相关计算资源后，请务必及时释放资源，否则会造成资源浪费甚至账号欠费。

ModelArts 以下模块会涉及云资源的消耗并会按需或按周期产生费用：自动学习训练和部署，Notebook 实例，训练作业，TensorBoard，在线服务。使用完需要“停止”自动学习部署、Notebook、TensorBoard、在线服务。

停止所有需消耗云资源的作业后，可在 ModelArts 总览页面确认所有计费项都为“0”。



ModelArts

总览

自动学习

数据标注Beta

数据管理

开发环境

训练作业

模型管理

部署上线

自动学习	
图像分类	物体检测
0 计费中	0 计费中
1 项目	2 项目
 exeML-butterfly	 exeML-yunbao

AI全流程开发				
数据管理	开发环境	训练作业	模型管理	部署上线
0 计费中	0 计费中	0 计费中	0 计费中	0 计费中
4 数据集	9 实例	7 / 1 作业版本/TensorBoard	19 模型	18 服务
dataset-flowers	 notebook-lijf	 trainjob-lijf	model-lijf	 service-lijf

ModelArts 详细计费说明可参见：https://support.huaweicloud.com/price-modelarts/modelarts_07_0002.html。

注意

- 公有云的按需使用是采用后付费模式，即先使用再付费，一个小时会出一次话单，比如：2019.6.24 04:00~05:00 这段时间使用的费用，在 05:00 时刻采集完成后，才会出话单，进行扣费。

A 附录

本 baseline 相关代码文件 src.zip 下载地址：https://modelarts-competitions.obs.cn-north-1.myhuaweicloud.com/garbage_classify/codes/tensorflow/src.zip。

src.zip 文件说明见下表。

名称	说明
run.py	模型训练启动文件（创建训练作业选择“启动文件”时，选择 run.py）。
models 文件夹&其他.py 文件	run.py 运行时的相关依赖代码文件。
deploy_scripts	包含模型配置文件 config.json，模型推理代码 customize_service.py。
readme.docx	代码说明文档。