

华为云 “云上先锋” · AI 学习赛 Baseline【完整版】

ModelArts 是面向 AI 开发者的一站式开发平台，您可以使用 ModelArts 完成垃圾分类学习赛[垃圾分类学习赛](#)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 完成垃圾分类学习赛的流程指导，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 来完成垃圾分类学习赛题目。

温馨提示：

- **ModelArts 使用过程中，部分步骤将会产生费用，建议在开始做比赛之前就领取代金券，以免欠费。如果在领取代金券之前，您的账号已欠费，则必须自行充值抵扣欠费，代金券不能抵扣已有的欠费。**
- **注册华为云账号、实名认证、报名参赛，这三个条件都满足后，才能在大赛的赛事介绍页面有代金券可以领取，点击后领取就可以，不需要额外操作就可以直接使用。**

1. 数据和 Baseline 代码准备

ModelArts 使用对象存储服务（OBS）存储数据和模型的备份，实现安全、高可靠和低成本存储需求。竞赛数据集已上传至华为云对象存储服务（OBS）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 中执行几行简单的代码，将竞赛数据集和 line 代码拷贝至您创建的 OBS 桶中。

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**、**代码文件**及模型训练输出的**模型文件**，具体操作如下：

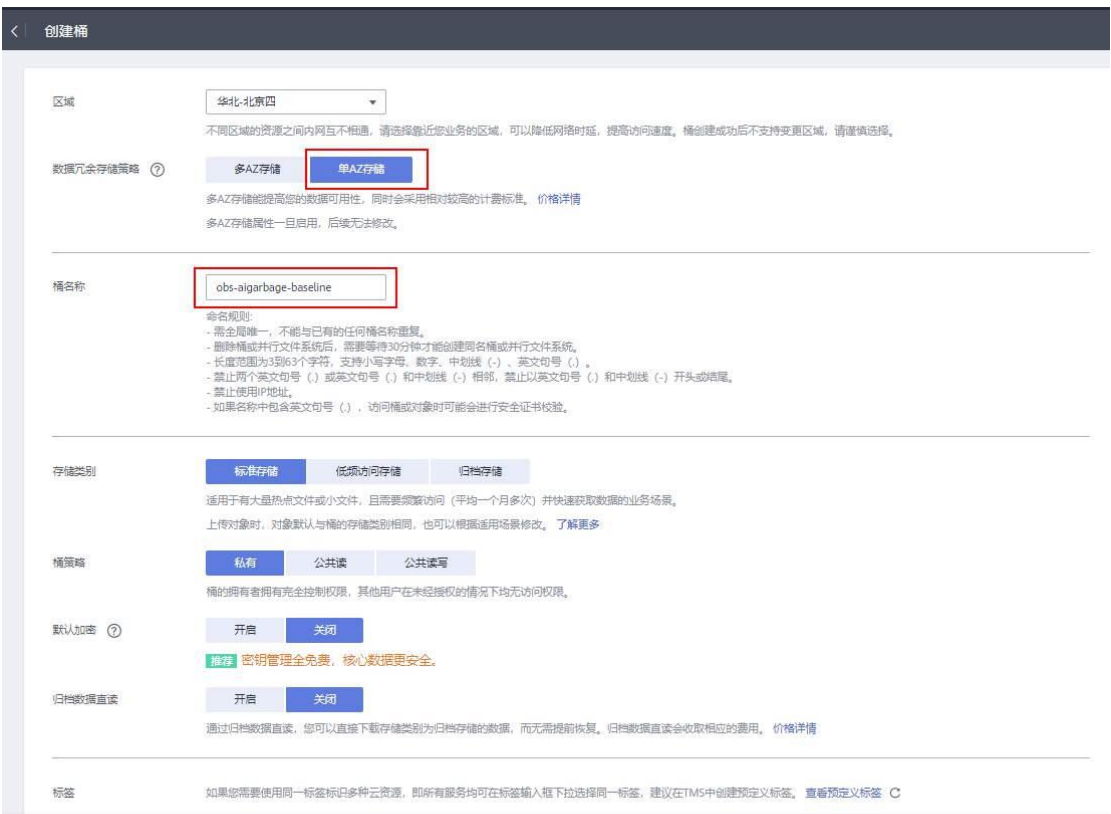
1. 登录 OBS 管理控制台 <https://storage.huaweicloud.com/obs/?®ion=cn-north-4&locale=zh-cn#/obs/buckets>。

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，单击“立即创建”。如图：

1) 创建桶



2) 自定义参数



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

华为云社区
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，不要使用数字开头。
- 建议使用 “单 AZ 存储”。

3. 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竞赛数据集；如：dataset
- 新建一个文件夹存放代码文件；如：codes
- 新建一个文件夹存放训练输出的模型文件； 如：model_output



创建 Notebook 并通过代码拷贝竞赛数据集和 line 代码

本节介绍如何在 ModelArts 开发环境创建 Notebook，并在 Notebook 中通过代码将竞赛数据集和代码拷贝至您的 OBS 桶中，具体操作如下：

1. 点击进入 [ModelArts 开发环境](#)。
2. 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。

创建 Notebook < 返回 Notebook 列表

1 服务选型 2 规格确认 3 完成 使用指南 建议反馈

* 计费模式 按需计费

* 名称 notebook-aigabage-baseline

描述

* 工作环境 公共镜像 我的镜像

名称	描述
<input checked="" type="radio"/> Multi-Engine 1.0 (python3 推荐)	预置AI引擎: MXNet-1.2.1, PySpark-2.3.2, Pytorch-1.0.0, TensorFlow-1.13.1, TensorFlow-1.8, XGBoost-Sklearn
<input type="radio"/> Multi-Engine 1.0 (python2)	预置AI引擎: Caffe-1.0.0, MXNet-1.2.1, PySpark-2.3.2, PyTorch1.0.0, TensorFlow-1.13.1, TensorFlow-1.8, XGBoost-Sklearn
<input type="radio"/> Multi-Engine 2.0 (python3)	预置AI引擎: Pytorch-1.4.0, R-3.6.1, TensorFlow-2.1.0
<input type="radio"/> Ascend-Powered-Engine 1.0 (python3)	预置AI引擎: MindSpore-0.5.0, TensorFlow-1.15.0

* 资源池 公共资源池 专属资源池

* 类型 CPU GPU

* 规格 限时免费)体验规格CPU版 2核 8GiB 8核 32GiB

适合场景: 适合在Notebook上的代码编写与体验

1、免费规格用于体验, 会在1小时后自动停止, 72小时内没有再次启动, 会释放资源, 请注意文件备份。每个用户只限创建一个基于此免费规格的实例。

2、ModelArts免费算力不包含对象存储服务(OBS)存储资源费用, 对象存储服务(OBS)计费标准详见如下链接: 对象存储服务(OBS)计费详情。

☒ 我已阅读并同意以上内容

存储配置 云硬盘(EVS) 对象存储服务(OBS)

Notebook文件管理页面显示/home/ma-user/work目录(云硬盘(EVS)挂载位置), 只有该目录下的数据在Notebook停止后不会被清理。

* 磁盘规格 5 GB 最小值: 5; 最大值: 4,096

1 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按照标准费用收费。

Git存储库

费用 ¥0.00/小时

下一步

注意： 为方便在 ModelArts Notebook 中调试代码，“存储配置”建议选择“EVS”。

3. 单击“下一步”进行规格确认，确认无误后单击“提交”。



4. 单击"返回 Notebook 列表", 在列表操作栏中点击 "打开", 打开 Notebook。

1) 返回 Notebook 列表



2) 打开 Notebook

名称	状态	工作环境	规格	描述	创建时间	创建者	操作
notebook-aigarbage-base...	运行中 (57 分钟后停止)	Python3 GPU	GPU: 1*100 CPU: 8 核 64GB	--	2020/11/13 09:26:01 GMT+08...	hu40161597	打开 打开JupyterLab 停止 删除

5.单击 "Files" 页签右上角 "New", 选择 "TensorFlow-1.13.1" 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和 line 代码拷贝至您创建的 OBS 桶中。

```
import moxing as mox

# 拷贝代码
mox.file.copy_parallel('s3://ma-competitions-bj4/waste_sorting/src', 's3://obs-aigarbage-baseline/codes/src')

# 拷贝训练数据
mox.file.copy_parallel('s3://ma-competitions-bj4/waste_sorting/datasets', 's3://obs-aigarbage-baseline/dataset/garbage_classify')

print('Copy procedure is completed !')
```

当打印出 “Copy procedure is completed ! ” 即为拷贝完成。

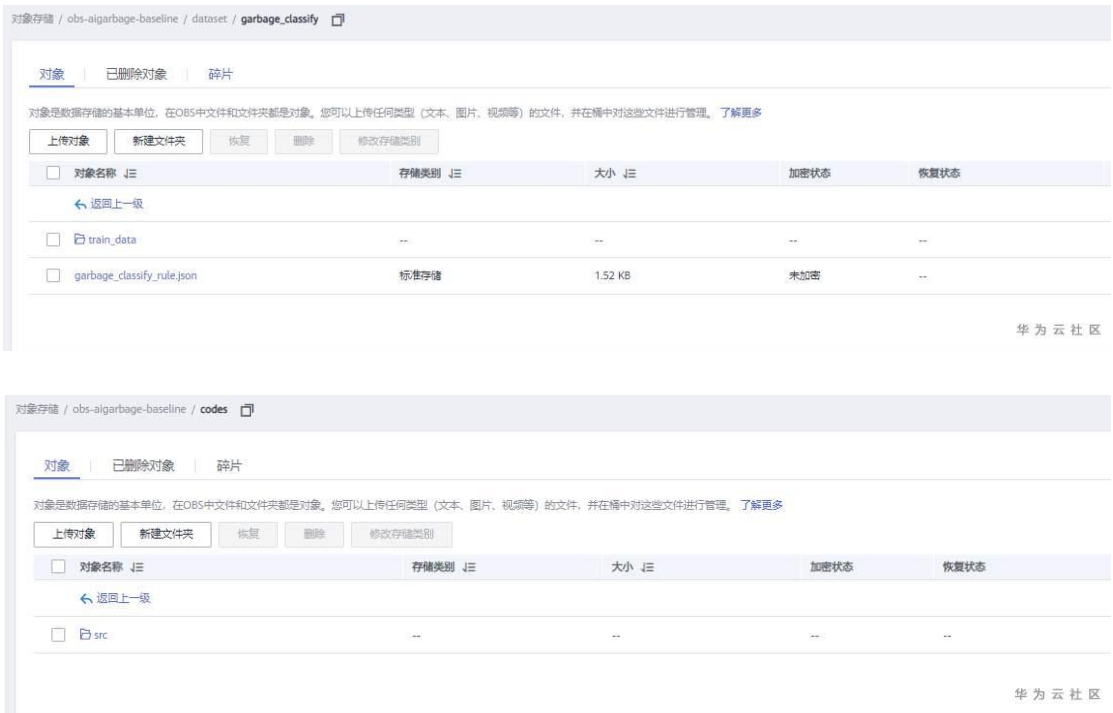
如图：



将 “obs-aigarbage-baseline” 替换为您创建的 OBS 桶名称。

- 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。

拷贝成功后，可在您的 OBS 桶中查看,如下是 Baseline 的 OBS 桶拷贝成功的截图。



2. 在 ModelArts 全局配置中配置访问授权

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 全局配置中配置访问授权。授权方式有两种，推荐使用委托方式授权。

使用委托授权（推荐）

- 1.使用华为云账号登录 [ModelArts 管理控制台](#)，在左侧导航栏单击“全局配置”，进入“全局配置”页面，单击“访问授权”。



2. 在弹出的“访问授权”窗口中，“授权方式”选择“使用委托”，选择需要授权的“用户名”及其对应的“委托”。点击“自动创建”按钮，使用 ModelArts 自动创建的委托名即可。
3. 完成“用户名”和“委托”配置后，勾选“我已经详细阅读并同意《ModelArts 服务声明》”，然后单击“同意授权”，完成授权访问。



3.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“训练管理”，单击“训练管理 > 训练作业 > 创建”按钮进入到“创建训练作业”配置页面。



2. 在创建训练作业页面填写参数，参数填写说明如下。

- 名称：

可按照自己的习惯命名。

- 算法来源：

选择“常用框架”，本 line “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。

- 数据来源：

选择“数据存储位置”，注意 OBS 路径要选择到 train_data。

- 训练输出位置：

OBS 路径选择到 model 文件夹上一层目录，本 line 为/obs-aigargbage-line /model_output。

- 运行参数：

新增 `deploy_script_path = s3://ai-competitions-bj4/waste-sorting/src/deploy_scripts`，推理代码&配置文件的路径，此处为固定路径，直接复制粘贴即可。

- 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。

- 资源池选择：公共资源池
- 类型：GPU
- 规格：GPU：1 * nvidia-v100 32GiB CPU：8 核 64GiB
- 计算节点个数：使用默认值 1。

填写完成后点击“下一步”。

创建训练作业

返回训练作业列表

1 服务选型

2 规格确认

3 完成

使用指南

建议反馈

* 计费模式

按需计费

* 名称

trainjob-garbage-competiton

版本

V0001 版本信息为自动生成

描述

0/256

一键式参数配置

如您已保存过参数配置，可单击 [这里](#)

* 算法来源

算法管理

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

/obs-aigarbage-baseline/codes/src/

选择

* 启动文件

/obs-aigarbage-baseline/codes/src/run.py

选择

* 数据来源

数据集

数据存储位置

* 数据存储位置

/obs-aigarbage-baseline/dataset/garbage_classify/train_u

选择

* 训练输出位置

/obs-aigarbage-baseline/model_output/

选择

请尽量选择空目录来作为训练输出路径。

华为云社区

运行参数

train_url

=

/obs-aigarbage-baseline/model_output/

data_url

=

/obs-aigarbage-baseline/dataset/garbage_u

deploy_script_path

=

s3://ai-competitions-bj4/waste-sorting/src/

增加运行参数

作业日志路径

/obs-aigarbage-baseline/logs/

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。

* 资源池

公共资源池

专属资源池

* 类型

CPU

GPU

* 规格

GPU: 1 * nvidia-v100 32GiB CPU: 8 核 64GiB

该规格当前用户数较多，作业需要排队。

* 计算节点个数

-

1

+

订阅消息

保存作业参数

配置费用 ¥28.00/小时

云社区

下一步

3. 信息确认无误，点击“提交”按钮。

创建训练作业

< 返回训练作业列表

1 服务选型

2 规格确认

3 完成

使用指南

建议反馈

规格确认

产品名称	产品规格	计费模式	价格
	作业名称	trainjob-garbage-competition	
	描述	--	
	训练数据集	/obs-aigarbage-baseline/dataset/garbage_classify/train_data/	
	AI引擎	TensorFlow TF-1.13.1-python3.6	
	训练输入	--	
	训练输出	--	
trainjob-garbage-competit...	代码目录	/obs-aigarbage-baseline/codes/src/	
	启动文件	/obs-aigarbage-baseline/codes/src/run.py	
	运行参数	deploy_script_path=s3://ai-competitions-bj4/waste-sorting/src/deploy_scripts	
	训练输出位置	/obs-aigarbage-baseline/model_output/	
	日志输出位置	/obs-aigarbage-baseline/logs/	
	规格	GPU: 1 * nvidia-v100 32GiB CPU: 8 核 64GiB	
	计算节点个数	1	

配置费用 ¥28.00/小时

华为云社区

上一步

提交

4. 点击“查看作业详情”按钮。

创建训练作业

< 返回训练作业列表

1 服务选型

2 规格确认

3 完成

使用指南

建议反馈

任务提交成功!

您的训练作业trainjob-garbage-competition已开始创建, 请耐心等待。

查看作业详情

返回训练作业列表

您的余额: ¥ 0.95

充值

华为云社区

进入作业详情页，等待状态为“运行完成”，即为训练成功。

训练作业 / trainjob-garbage-competition

版本管理

版本列表 版本对比

2020/11/13 10:08:57 当前版本: V0001 状态: **运行中** 运行时间: 00:01:41 C

创建可视化作业 创建模型 修改 更多操作

配置信息	日志	资源占用情况	评估结果
作业名称	trainjob-garbage-competition job5cfeef		
状态	运行中		
运行版本	V0001		
开始运行时间	2020/11/13 10:09:06 GMT+08:00		
运行时间	00:01:41		
规格	CPU: 8 核 64GB GPU: 1 * nvidia-v100 32GB		
计算节点个数	1		
日志输出位置	/job-aigarbage-baseline/logs/		
NAS 挂载路径	-		
AI引擎	TensorFlow / TF-1.13.1-python3.6		
代码目录	/job-aigarbage-baseline/codes/src/		
启动文件	/job-aigarbage-baseline/codes/src/run.py		
训练数据集	/job-aigarbage-baseline/dataset/garbage_classify/train_data/		
运行参数	deploy_script_path=/ma-competitions-44/waste_sorting/src/deploy_scripts		
训练输出位置	/job-aigarbage-baseline/model_output/		
描述	-		
NAS 地址	-		

华为云社区

此训练过程需运行 5 分钟左右，请耐心等待。

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 控制台](#)，单击左侧导航栏选择“模型管理 > 模型”，单击页面中

“导入”按钮。

ModelArts

总览

活动

自动学习

数据管理

开发环境

算法管理

训练管理

模型管理

模型

压缩/转换

评估/诊断

部署上线

AI市场

专属资源池

全局配置

模型

我的模型 市场订阅模型 云服务订阅模型

导入 查找模型

模型名称	最新版本	部署类型	版本数量
AI-final	0.0.1	在线服务/批量服务/边缘服务	1
ai-2	0.0.1	在线服务/批量服务/边缘服务	1
ai	0.0.2	在线服务/批量服务/边缘服务	2
ai-1	0.0.1	在线服务/批量服务/边缘服务	1
ai-0	0.0.1	在线服务/批量服务/边缘服务	1
ai-fcda41c	0.0.9	在线服务/批量服务	1
ai-terRCNN	0.0.1	在线服务	1
ai-0.1	0.0.1	在线服务	1
ai-0.1	0.0.1	在线服务	1

10 总条数: 28 1 2 3 跳转 1

华为云社区

2.在导入模型页面填写：

- 名称：自定义。
- 版本：使用缺省值。
- 标签，描述：可根据需要填写。
- 元模型来源：从 OBS 中选择，请选择 model 文件夹上一级目录；本 line model 文件夹上一级目录为/obs-aigarbage-line/model_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。

选择好参数后点击“立即创建”，进行模型导入。

导入模型

返回模型

使用指南

建议反馈

名称

model-garbage-competition

版本

0.0.1

标签

输入标签，按Enter键创建

描述

0/100

元模型来源

从训练中选择

从模型中选择

从容器镜像中选择

从对象存储服务 (OBS) 中选择

此选项将导入您在对象存储服务 (OBS) 中的模型，目前仅支持从对象存储服务 (OBS) 导入TensorFlow、MXNet、Caffe、Spark MLlib、Scikit Learn、XGBoost、PyTorch、Image8 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，选择模型时请您选择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

/obs-aigarbage-baseline/model_outpu

AI引擎

TensorFlow

部署类型

☒ 在线服务

☐ 批处理服务

☐ 边缘服务

配置文件

☐

推理代码

https://obs-aigarbage-baseline.obs.myhwclouds.com/model_output/model/customize_sa

华为云社区



3. 等待模型导入完成，大约需要 3 分钟（状态显示为“正常”），点击左侧下拉箭头，可以查看导入状态。

1) 构建中



2) 导入完成



5.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。

部署为在线服务具体步骤如下。

1.在导入完成的模型中，点击右侧的“部署 > 在线服务”。

我的模型

市场订阅模型

云市场订阅模型

导入

克隆模型

全部模型

请输入名称和模型

模型名称	最新版本	部署类型	版本数量	创建时间	描述	操作	
model-garbage-competition	0.0.1	在线服务	1	2020/11/13 10:31:10 GMT+08:00		<div>创建新版本</div> <div>删除</div>	
<div>请输入版本名称</div>							
版本	状态	部署类型	模型大小	模型来源	创建时间	描述	操作
0.0.1	<div>正常</div>	在线服务	141.44 MB	自定义算法	2020/11/13 10:31:10 GMT+08:00		<div>部署</div> <div>发布</div> <div>删除</div>
model-5d96	0.0.1	在线服务/批量服务/边缘服务			2020/11/12 21:27:38 GMT+08:00		<div>在线服务</div> <div>批量服务</div> <div>边缘服务</div> <div>创建新版本</div> <div>删除</div>

华为云社区

2.在部署页面填写参数:

- 名称：自定义。
- 是否自动停止：因为在线服务会持续收费，如开启了自动停止，则在指定时间到了后会自行停止（建议选择“1 小时后”，避免产生不必要的费用）。
- 资源池：公共资源池。
- 模型来源选择前面步骤导入的模型。
- 计算节点规格：CPU：2 核 8 GiB

部署
< 返回在线服务

1 服务选型
2 规格确认
3 完成

使用指南
建议反馈

* 计费模式
按需计费

* 名称
service-garbage-competition

是否自动停止
☒

开启该选项后，在线服务的运行时间将在您选择的时间点后，自动停止，同时服务计费停止

☒ 1小时后
☐ 2小时后
☐ 4小时后
☐ 6小时后
☐ 自定义

描述

0/100

* 资源池
公共资源池
专属资源池

* 选择模型及配置

模型来源
我的模型
市场订阅模型

模型
model-garbage-competition
0.0.1
分流 (%)
100

计算节点规格
CPU: 2 核 8 GiB
计算节点个数
1

适合场景：CPU标准规格，满足大多数模型的运行和预测

环境变量
增加环境变量

华为云社区

数据采集 ①

准测筛选 ①

支持APP认证 ①

订阅消息 ①

配置费用 ¥0.80/小时 ②

下一步

3.单击“下一步”，参数确认无误后，单击“提交”。

部署

① 服务选型 ② 规格确认 ③ 完成

使用指南 | 建议反馈

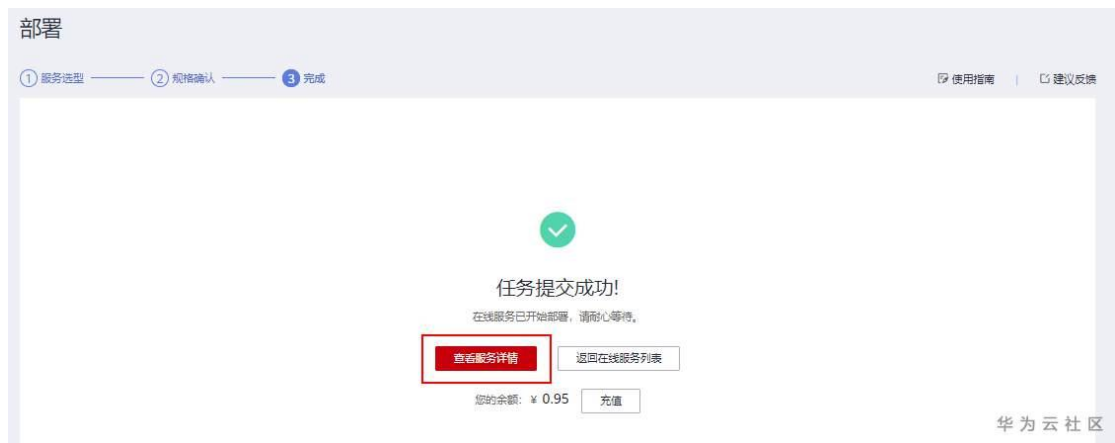
详情

产品类型	产品规格		计费模式	数量	价格
在线服务	名称	service-garbage-competition	按需计费	1	¥0.80/小时
	自动停止时间	1 小时			
	模型名称	model-garbage-competition			
	模型版本	0.0.1			
	分流 (%)	100 %			
	计算节点规格	CPU: 2 核 8 GiB			
	计算节点个数	1			
	环境变量	--			
参数配置					

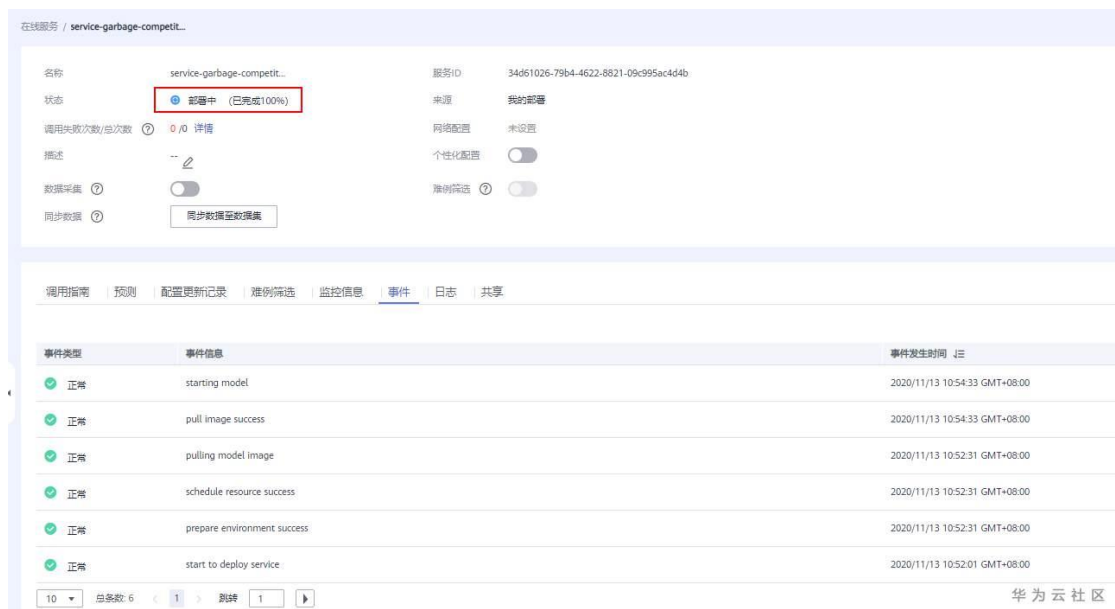
配置费用 ¥0.80/小时 ②

上一步 提交

4.点击“查看服务详情”。



5.查看部署进度。



6.部署完成，点击“预测”进入预测页签，从本地上传一张图片，点击“预测”按钮进行预测。

在线服务 / service-garbage-competit...

使用指南 | 建议反馈 | 修改

名称service-garbage-competit...

服务ID34d51026-79b4-4622-8821-09c95ac4d4b

状态运行中 (50 分钟 后停止)

来源我的部署

调用失败次数/总次数0/0 详情

网络配置未设置

描述

个性化配置

数据采集中

预测延迟


同步数据同步数据至数据湖

调用指南 | 预测 | 配置更新记录 | 推理筛选 | 监控信息 | 事件 | 日志 | 共享

请求路径: / 选择预测文件 上传 预测 推理区域

预测图片预览

预测结果显示



1

华为云社区

7. 预测结果如下:

使用指南 | 预测 | 配置更新记录 | 推理筛选 | 监控信息 | 事件 | 日志 | 共享

请求路径: / 选择预测文件 上传 预测 推理区域

预测图片预览

预测结果显示



预测成功

```
1 {
2   "result": "其他垃圾/一次性纸制品"
3 }
```

华为云社区

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

华为云社区

3) 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、反馈信息。其中“提交作品”页面需报名比赛后才会显示。排行榜每 6 小时更新一次。

6.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 详细计费说明可参见：https://support.huaweicloud.com/productdesc-modelarts/modelarts_01_0021.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。