

## 使用 MoXing 实现冰山图像分类应用

本文介绍如何在华为云 ModelArts 平台上使用 MoXing 实现 Kaggle 竞赛中的冰山图像分类任务。实验所使用的图像为雷达图像，需要参赛者利用算法识别出图像中是冰山（iceberg）还是船（ship）。操作的流程分为 4 部分，分别是：

1. **准备数据**：下载数据集并上传至华为云 OBS 桶中。
2. **训练模型**：使用 MoXing API 编写用实现冰山图像分类的网络模型，新建训练作业进行模型训练。
3. **预测结果**：再次新建训练作业，对 test 数据集进行预测，并将结果保存到 csv 文件。
4. **查看结果**：将预测结果的 csv 文件提交到 Kaggle 官网后获取分类结果。

### 1. 准备数据

**步骤 1** 首先下载冰山图像分类数据集 iceberg（数据集格式为 TFRecord），数据集为如下三个文件，点击超链接进行下载：

- [iceberg-train-1176.tfrecord](#)：训练数据集
- [iceberg-eval-295.tfrecord](#)：验证数据集
- [iceberg-test-8424.tfrecord](#)：预测数据集

**步骤 2** 参考“[上传业务数据](#)”章节内容，数据集上传至华为云 OBS 桶（假设 OBS 桶路径为：s3://obs-iceberg-test/dataset/）。

**步骤 3** 登录“[ModelArts](#)”管理控制台，在“全局配置”界面添加访问密钥。

图 1 添加访问密钥



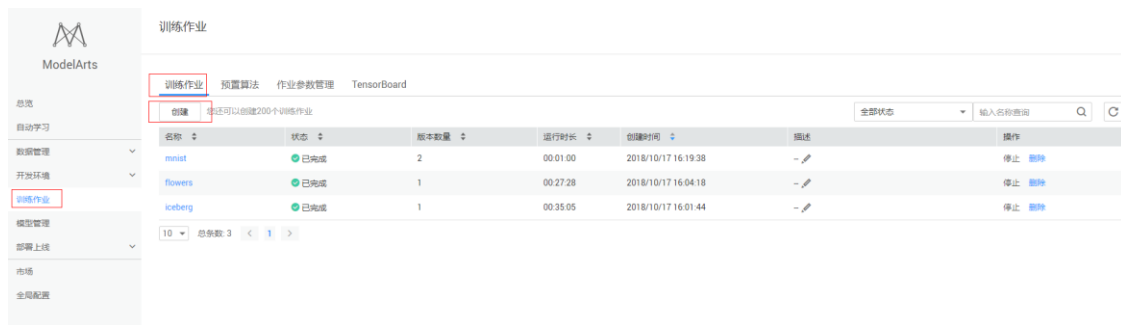
## 2. 训练模型

将模型训练脚本上传至 OBS 桶中（您也可以在 ModelArts 的开发环境中编写模型训练脚本，并转成 py 文件），然后创建训练作业进行模型训练，操作步骤如下：

**步骤 1** 参考“[上传业务数据](#)”章节内容，将模型训练脚本文件 [train\\_iceberg.py](#) 上传至华为云 OBS 桶（假设 OBS 桶路径为：s3://obs-iceberg-test/code/）。

**步骤 2** 返回“ModelArts”控制台，在“训练作业”界面。单击左上角的“创建”。

图 2 训练作业界面



**步骤 3** 填写参数。“名称”和“描述”可以随意填写，“数据来源”请选择“数据的存储位置”（s3://obs-iceberg-test/dataset/），“算法来源”请选择“常用框架”，“AI引擎”选择“TensorFlow”(TF-1.8.0-python3.6)，“代码目录”请选择训练脚本文件 train\_iceberg.py 所在的 OBS 父目录（s3://obs-iceberg-test/code/），“启动文件”请选择“train\_iceberg.py”，“训练输出位置”请选择一个路径（建议新建一个文件夹）用于保存输出模型和预测文件（例如：s3://obs-iceberg-test/log/），参数确认无误后，单击“立即创建”，完成训练作业创建。

图 3 训练作业参数配置

创建训练作业 < 返回作业列表

\* 名称

train\_iceberg

版本

V0001

版本信息为自动生成

描述

0/256

一键式参数配置

如果您创建之前已保存过参数配置, 可选择已有参数帮助您快速配置, 点击选择 [作业参数配置](#)。

\* 数据来源

数据集

数据存储位置

\* 数据存储位置

/obs-iceberg-test/dataset/

选择

\* 算法来源

预置算法

常用框架

\* AI引擎

TensorFlow

TF-1.8.0-py...

\* 代码目录

/obs-iceberg-test/code/

选择

\* 启动文件

/obs-iceberg-test/code/train\_iceberg.py

选择

运行参数

+ 增加运行参数

\* 训练输出位置

/obs-iceberg-test/log/

选择

一般训练输出位置为空目录, 如果该目录下已有文件, 请确保这些文件需要被加载。

\* 计算节点规格

2核 | 8GiB

2核 | 8GiB | 1\*P100

\* 计算节点个数

-

1

+

**步骤 4** 在模型训练的过程中或者完成后, 可以通过创建 TensorBoard 作业查看一些参数的统计信息, 如 loss, accuracy 等。在“训练作业”界面, 点击 TensorBoard, 再点击“创建”按钮, 参数“名称”可随意填写, “日志路径”请选择步骤 3 中“训练输出位置”参数中的路径 ( s3://obs-iceberg-test/log/ )。

图 4 创建 tensorboard



训练作业完成后，即完成了模型训练过程。如有问题，可点击作业名称，进入作业详情界面查看训练作业日志信息。

### 3. 预测结果

待训练作业运行完成后，在“训练输出位置”目录下生成模型文件（如：`model.ckpt-5600`）。由于我们只需要进行一次预测，因此不需要部署在线预测服务。相关的预测操作已经在“train\_iceberg.py”文件写好，预测结果将输出到“submission.csv”文件。我们使用训练作业进行预测，操作步骤如下：

**步骤 1** 参考“2.训练模型”在“训练作业”界面，单击左上角的“创建”。

**步骤 2** 填写参数，参考“2.训练模型 步骤 3”填写参数，“训练输出位置”请保持一致（`s3://obs-iceberg-test/log/`），预测‘计算节点个数’只能选择 1 个节点，另外添加参数“`is_training=False`”单击“立即创建”，完成训练作业创建。

创建训练作业 < 返回作业列表

\* 名称

predict\_iceberg

版本

V0001

版本信息为自动生成

描述

0/256

一键式参数配置

如果您创建之前已保存过参数配置,可选择已有参数帮助您快速配置,点击选择 [作业参数配置](#)。

\* 数据来源

数据集

数据存储位置

\* 数据存储位置

/obs-iceberg-test/dataset/

选择

\* 算法来源

预置算法

常用框架

\* AI引擎

TensorFlow

TF-1.8.0-py...

\* 代码目录

/obs-iceberg-test/code/

选择

\* 启动文件

/obs-iceberg-test/code/train\_iceberg.py

选择

运行参数

is\_training

=

False

✕

+ 增加运行参数

\* 训练输出位置

/obs-iceberg-test/log/

选择

一般训练输出位置为空目录,如果该目录下已有文件,请确保这些文件需要被加载。

\* 计算节点规格

2核 | 8GiB

2核 | 8GiB | 1\*P100

\* 计算节点个数

-

1

+

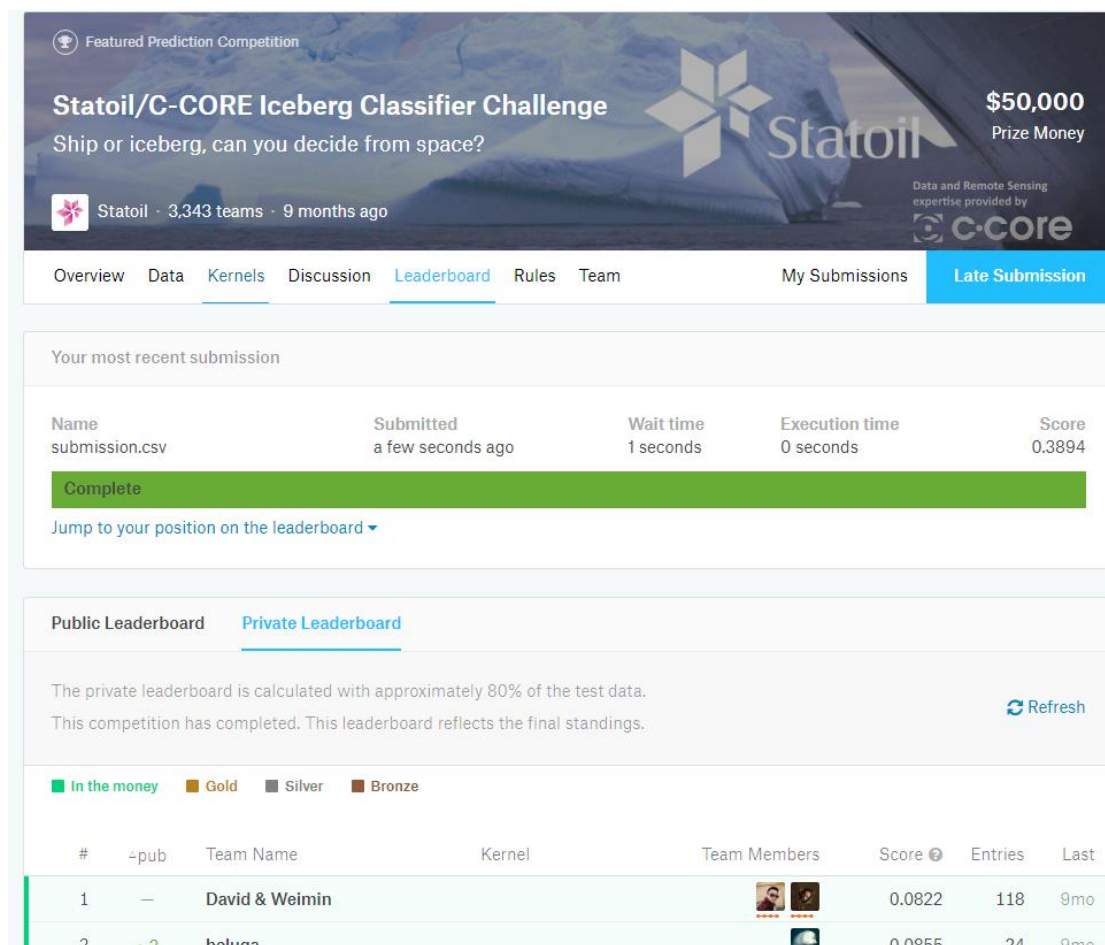
**步骤 3** 训练作业执行完成后,在“训练作业”界面,单击 predict\_iceberg 作业名称,进入作业的详情界面。在“训练日志”中,可以查看到在 eval 数据集上的 loss 值。在“训练输出位置”目录下 (s3://obs-iceberg-test/log/), 能看到用于保存预测结果的“submission.csv”文件。

## 4. 提交预测结果

登录 [Kaggle 官网](#)，注册并登录账号。（注册时若中国大陆地区手机号是：

18688888888，手机号填写格式应为：+86018688888888），

将“submission.csv”文件提交到 Kaggle 冰山识别界面“[Statoil/C-CORE Iceberg Classifier Challenge](#)”，点击右上角的 Late Submission 上传文件得到预测结果的准确率（错误率）。



Featured Prediction Competition

### Statoil/C-CORE Iceberg Classifier Challenge

Ship or iceberg, can you decide from space?

Statoil · 3,343 teams · 9 months ago

\$50,000 Prize Money

Data and Remote Sensing expertise provided by c-core

Overview Data Kernels Discussion **Leaderboard** Rules Team My Submissions **Late Submission**

Your most recent submission

Name	Submitted	Wait time	Execution time	Score
submission.csv	a few seconds ago	1 seconds	0 seconds	0.3894

Complete

[Jump to your position on the leaderboard](#)

Public Leaderboard **Private Leaderboard**

The private leaderboard is calculated with approximately 80% of the test data.  
This competition has completed. This leaderboard reflects the final standings.

Refresh

■ In the money ■ Gold ■ Silver ■ Bronze

#	pub	Team Name	Kernel	Team Members	Score	Entries	Last
1	—	David & Weimin			0.0822	118	9mo
2	▲3	beluga			0.0855	24	9mo

## 5. 打卡任务

熟悉ModelArts各部分功能，在ModelArts动手实践：

**创建 Notebook->执行数据转换，创建训练作业 →训练完成，创建训练作业  
(预测) →预测完成→ 删除训练作业。**

截图反馈训练完成的截图，需包含用户名（截图右上角），例如如下：

