

本作业分为两个难度，体验作业（低难度）和进阶作业（中等难度），推荐大家先完成体验作业，再尝试进阶作业

体验作业：完成 Evaluation 流程

体验作业提供：finetune 完毕的 tnews ckpt 文件、tnews 数据集、源代码

体验作业流程：

1. 申请华为云账号、申请 ModelArts 华为云昇腾集群服务公测
2. 从链接下载 tnews ckpt 文件、tnews 数据集和源代码
3. 创建 obs 桶
4. 上传 tnews ckpt 文件、tnews 数据集和源代码到 obs 桶
5. 提交 ModelArts 训练作业
6. 等待、查看结果

体验作业详细步骤

1. 申请华为云账号、申请 ModelArts 华为云昇腾集群服务公测



<https://www.huaweicloud.com/>

注册完毕之后，进入到 ModelArts 界面，点击昇腾集群公测，完成申请流程

AI开发平台ModelArts

ModelArts是面向开发者的一站式AI开发平台，为机器学习与深度学习提供海量数据预处理及半自动化标注、大规模分布式Training、自动化模型生成，及端-边-云模型按需部署能力，帮助用户快速创建和部署模型，管理全周期AI工作流。

按需/包周期付费可选，最低¥0.00/小时

购买套餐

进入控制台

ModelArts Pro

昇腾集群公测

AI开发者社区

ModelArts邀请好友 好礼1+1：AI算力 华为手表手环 周边纪念品随你拿 **福利**

<https://www.huaweicloud.com/product/modelarts.html>

2. 完成申请之后，开始下载所需文件，包括数据集文件、ckpt文件和源代码文件

tnews 数据集：

https://e-share.obs-website.cn-north-1.myhuaweicloud.com?token=SvjC1tq+9EC0fLT0WI5w1SIJKMVsoAn6d3hiNI5bHpGQM_MKuJiZhGfVhyPTgqUIFIVIYKtAmENUpN6ssH654op9P9KNXqX021meACXkf2EZ+O58B+FMoSmBJoai+97k52R89dhiOtO4HUMuIK6JYdZImjBOaZ1gxeB+gXCgHBKRNpzfcccU4d41LIA3BPBLqhJ24K0Q76EZdwPUFdf+oRgVqr9d4n01yNmZ9PIIYQILFSxfQGelojXp5jif1EEISBUvsvSm880An8VEAqYCMdiLcyfVE9rv0YFodid0EPzGmngOMa5p2Q/P6q3jJsjaPsElkX4SMpQ80v4DkJ4J+un4gISVfwf5VWlxObY7CbnP2YIJ+K7F+m9OrRN8oDm4aUooUfi0u8xIKlij4NqFyWMjUuPCsAIHavtpB+WQqMFOZooFxE45VAIZSXQiWIL3UKYVSm/Eyvi+cyg8PON0I501QJkhONQH7inkdoUMinwQoWOx/lyBzKiw8f94rtRtCWWNwl9DZsz/tDSF+yeMo3hI7p7NILWiFUbvFK2IUc=

提取码:

123456

bert 源代码：

<https://21days-bert.obs.cn-north-4.myhuaweicloud.com:443/bert.zip?AccessKeyId=M7KX8KLMT0ZL1P8QWXZ5&Expires=1634522579&Signature=LY3ZdN9gtoxx8ULDQRkkPuOp1nk%3D>

源代码下载完成后，在本地解压缩

tnews ckpt:

https://21days-bert.obs.myhuaweicloud.com:443/tnews-3_3335.ckpt?AccessKeyId=M7KX8KLMT0ZL1P8QWXZ5&Expires=1634522122&Signature=d21XsEKkzByK04oyZJ2c9JMGNLk%3D

这个文件比较大，请大家耐心下载

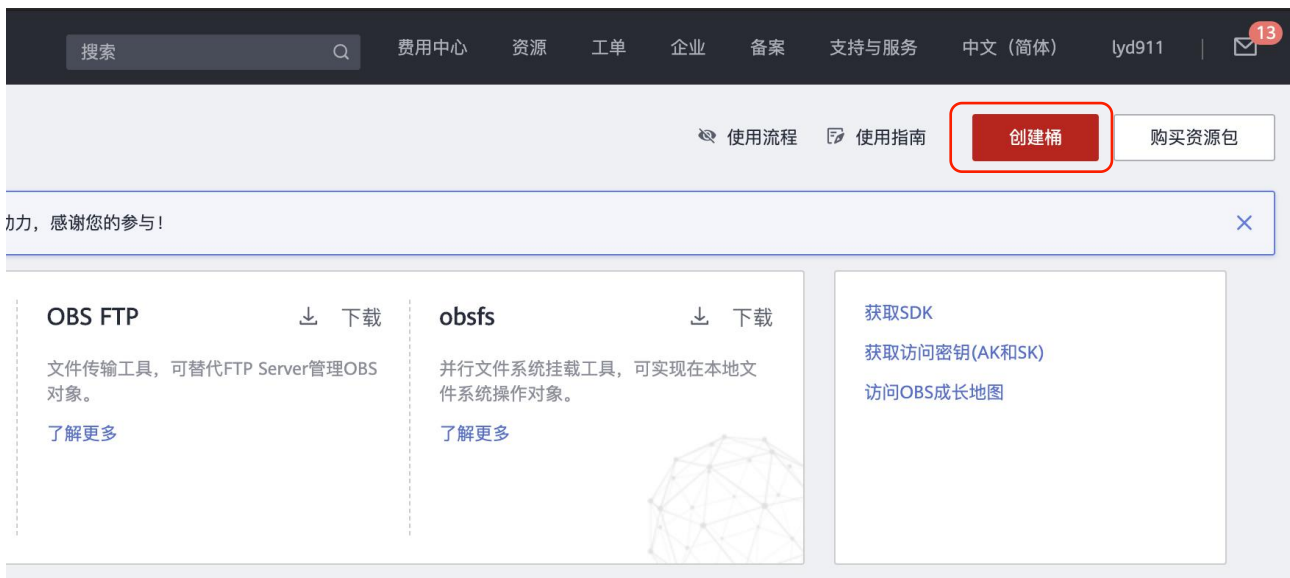
3. 创建 OBS 桶，进入官网界面 <https://www.huaweicloud.com/product/obs.html>，点击右上角的控制台按钮



在控制台界面中，在左上角选择北京四节点，同时在左边的弹出式菜单的搜索框中，搜索 OBS，并选择“对象存储服务 OBS”



此时我们已经进入了 OBS 对象存储服务的界面，点击右上角“创建桶”按钮



按照以下模板创建桶

数据冗余存储策略 ?

多AZ存储

单AZ存储

多AZ存储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](#)

多AZ存储属性一旦启用，后续无法修改。

桶名称

bert-21days

❗ 请求的桶名已经存在，或被其他用户占用，请重新输入。

命名规则:

- 需全局唯一，不能与已有的任何桶名称重复。
- 删除桶或并行文件系统后，需要等待30分钟才能创建同名桶或并行文件系统。
- 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

任何用户都可以对桶内对象进行读操作，仅桶拥有者可以进行写操作。

默认加密 ?

开启

关闭

推荐 密钥管理全免费，核心数据更安全。

归档数据直读

开启

关闭

数据冗余存储策略选择“单AZ存储”

桶名称自己取一个不重复的名称即可

桶策略采用公共读

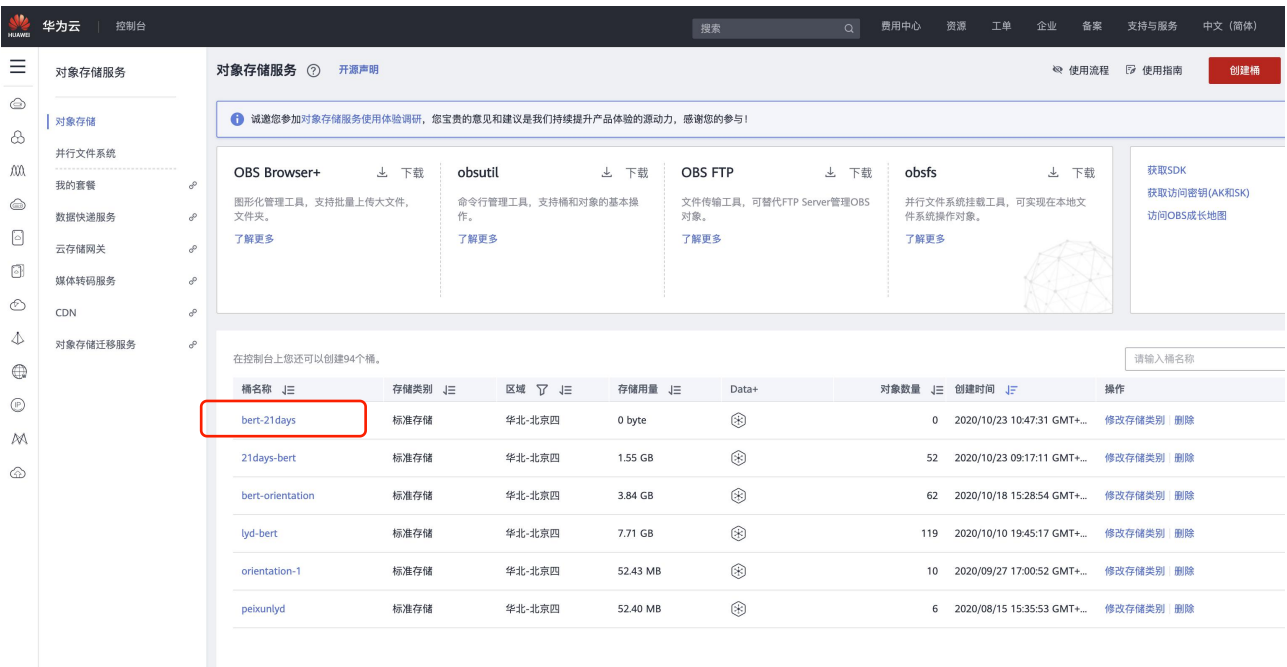
其他单策略使用默认单即可

点击创建按钮之后，OBS 桶就创建完毕了

4. 上传文件

这一步我们需要在刚刚创建的 OBS 桶内新建一些文件夹，并且上传我们刚刚下载好的文件

在 OBS 主页面点击进入刚才创建出来的桶



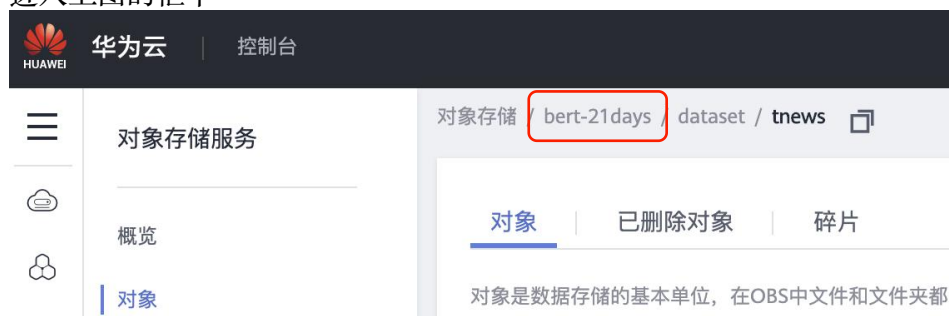
在左侧的弹出菜单中，点击“对象”，然后在右边点击新建文件夹,在根目录下新建两个文件夹：bert 和 dataset



点击进入 **dataset** 文件夹，再创建一个二级文件夹 **tnews**，并且进入 **tnews** 文件夹，点击“上传对象”按钮开始数据集的上传，可以采用拖拽的方式，直接将文件拖拽进入下图的红框中

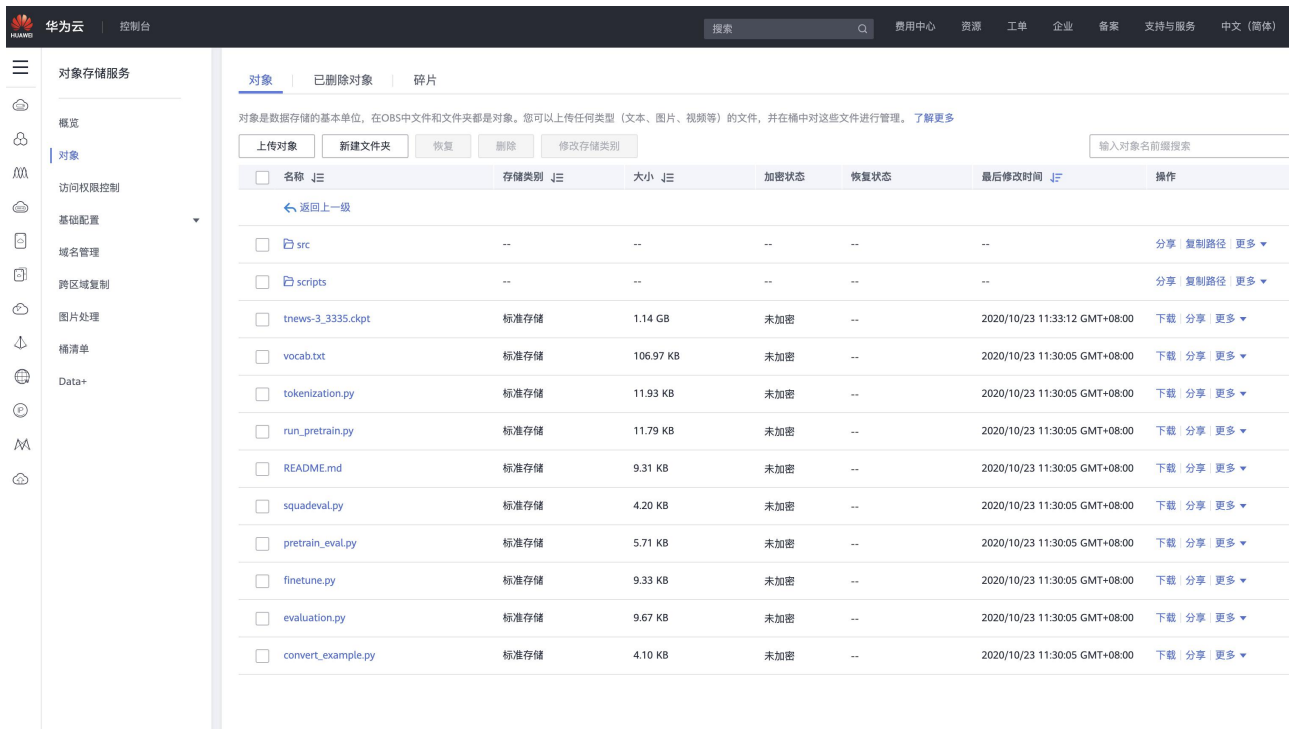


在完成了数据集的上传之后，回到桶的根目录，并且点击进入 **bert** 文件夹，同样的方法，开始上传解压缩之后的源代码文件，这里推荐使用拖拽的方式，把解压之后的文件和文件夹全选，然后拖拽进入上图的框中



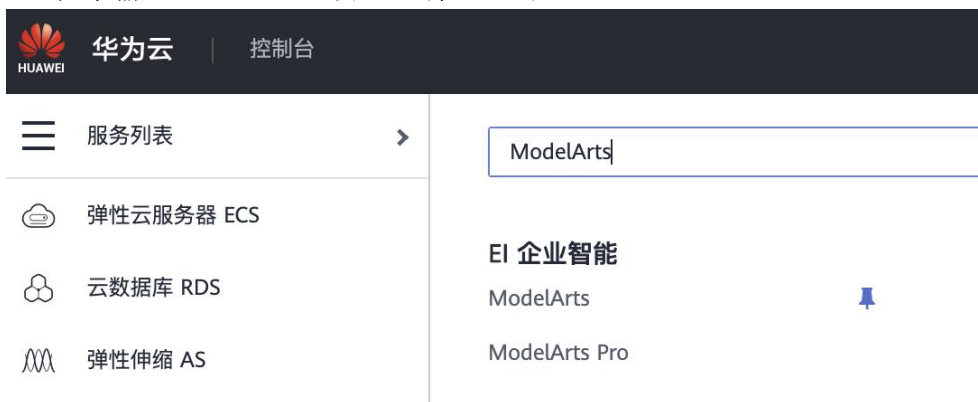
最后上传 **tnews ckpt** 文件，这里我们直接把 **tnews ckpt** 文件上传到 **bert** 目录即代码根目录下，由于 **tnews ckpt** 这个文件比较大，超过 1GB，所以上传时间会略微长一些

上传全部完成之后，截图如下：



至此，上传部分结束

5. 提交 ModelArts 训练作业，类似我们进入 OBS 的方式，在左侧服务列表的弹出菜单中，在搜索框中输入 modelarts，并且选择下面的“ModelArts”



进入 ModelArts 界面之后，点击左侧“训练管理”，在下拉菜单中点击“训练作业”，进入训练作业界面后，点击中上方的“创建”按钮



在创建训练任务的界面中，名称同样的，自己取一个不重复的名称即可

★ 名称 ✓

下面的参数配置中，以下几个参数需要配置：

“算法来源”选择“常用框架”，

“AI 引擎”选择“Ascend-Powered-Engine”和“MindSpore-0.5-python3.7”

“代码目录”选择之前在 obs 建立的 bert 目录即代码根目录

“启动文件”选择代码根目录下的 evaluation.py

“数据来源”选择“数据存储位置”

在“数据存储位置”中之前在 obs 中建立的 dataset 文件夹下的 tnews 文件夹

“训练输出位置”可以选择在 obs 根目录下新建一个文件夹“output”专门用来存储输出

“作业日志路径”跟“训练输出位置”类似，可以在 obs 根目录下创建一个“log”文件夹用来存放日志输出

“资源池”选择“公共资源池”，“类型”选择“Ascend”，“规格”选择“Ascend: 1 * Ascend 910 CPU: 24 核 96GiB”

其他参数选择默认即可

最终参数设置结果如下图所示：

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

Ascend-Powered-Engine

MindSpore-0.5-python3.7-aarch...

★ 代码目录

?

/bert-21days/bert/

选择

★ 启动文件

?

/bert-21days/bert/evaluation.py

选择

★ 数据来源

?

数据集

数据存储位置

★ 数据存储位置

/bert-21days/dataset/tnews/

选择

★ 训练输出位置

?

/bert-21days/output/

选择

请尽量选择空目录来作为训练输出路径。

运行参数

?

train_url

=

/bert-21days/output/

data_url

=

/bert-21days/dataset/tnews/

+ 增加运行参数

作业日志路径

?

/bert-21days/log/

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。

然后就可以提交任务了

6. 等待、查看结果

本身的 **evaluation** 任务时间是很短的，但如果同时提交任务但人数很多，可能会导致大部分人必须排队等待，等待的时间可长可短，所以到了这一步，大家也可以做一些其他事情，过一段时间回到 **ModelArts** 网页上检查一下作业状态即可。

如果显示“运行成功”，即代表 **Evaluation** 运行作业已经完成，我们可以，可以点击作业名称，进入到作业内部详情，再点击“日志”，查看输出

[版本管理](#)

版本过滤

版本对比

2020/10/23 12:55:44 当前版本: V0001 状态: ✔ 运行成功 运行时间: 00:03:09

配置信息

日志

资源占用情况

评估结果

作业名称

trainjob-bert-evaluation | joba73e8c60

状态

运行成功

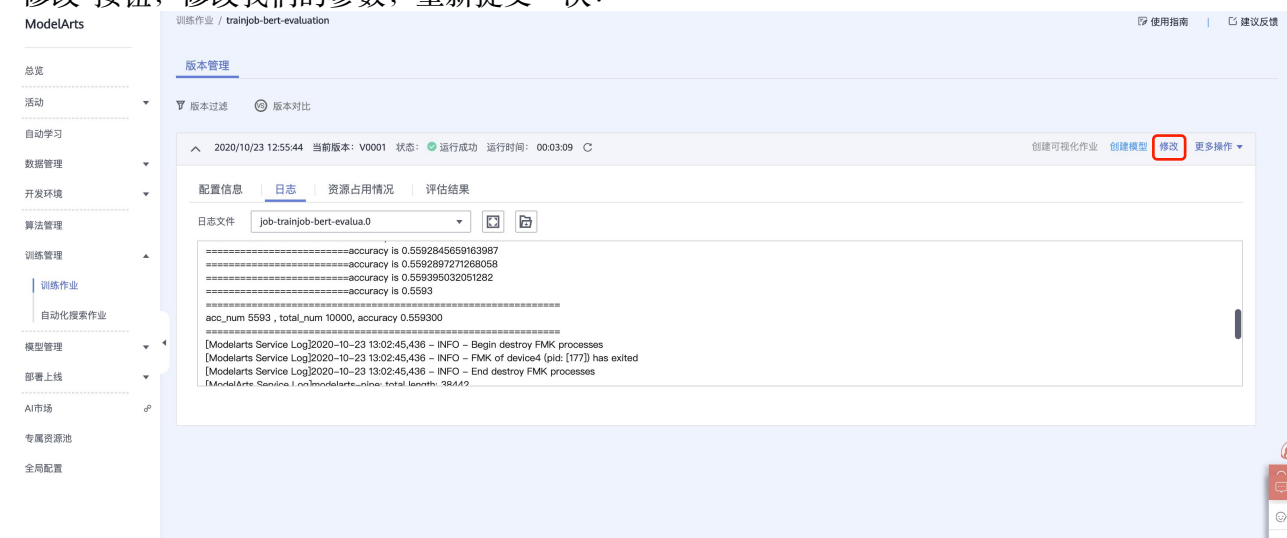
在日志里上下拖动滑动条，找到 accuracy 这一项，最后但这一条 accuracy 就是整个验证集但正确率



这样把正确率但截图上传完成了这次的体验作业

附加题：如果我们并不是想对整个验证集做推理，而只是想对一句话（一句中文短新闻）做推理，应该怎么做呢？

其实 1~4 步都是一样的，只有第 5 步略微不一样，我们可以回到提交过的作业界面，点击右上方的“修改”按钮，修改我们的参数，重新提交一次：



绝大多数参数都一样，只有一个地方有区别，在运行参数里面，我们点击下面的加号按钮，添加一行参数：

在左边的框中，输入“predict”，而在右边的参数框中，大家可以自由发挥输入一句中文的短新闻，我这里输入的是“星际争霸凉了”：

其他参数保持不变，提交作业

同样等待作业完成之后，查看日志：

可以看到分类显示“新闻_游戏”，到这一步，单句新闻的分类也完成了，大家可以截图上传作业