

《机器学习》第 11 周实验课流程

2022 年 4 月 26 日

一、Openbayes 实验环境搭建

链接：<https://openbayes.com/docs/getting-started/>

1. 注册 openbayes
网址 <https://openbayes.com>
2. 在页面左侧选择模型训练



3. 在界面中选择如图所示环境(NVIDIA VGPU, PyTorch-1.3.0-gpu)

新建训练: 基本信息 ☒ > 数据绑定 ☒ > 接入方式 ☒ > 审核

容器名称
输入一个容器名称, 可包含中文、英文、数字、连字符、下划线、半角句号, 最多 100 个字符

执行备注
输入执行备注, 用于备忘、描述此次执行的用途

选择算力
算力定义了 OpenBayes 的容器执行所提供的资源多少。 [了解更多](#)

cpu 无显卡 7 小时 27 分钟可用	NVIDIA T4 17.18 GB 显存 4 小时 10 分钟可用	NVIDIA Tesla V100 16GB 17.18 GB 显存 无可使用额度, 购买
cpu	t4	v100
NVIDIA VGPU 8.59 GB 显存 2 小时 54 分钟可用		
vgpu		

选择镜像
镜像提供了容器运行时所依赖的主要的机器学习算法库的种类。 [了解更多](#)

PyTorch 1.3.1-gpu	TensorFlow 1.15-gpu	OpenBayes automl-det-pytorch:1.11.0-py30-cu113
MXNet 1.6.0-gpu		

依赖管理
当前镜像中所安装的依赖包, 如果您需要安装自定义依赖, 请参考我们的依赖安装文档。如果你想安装额外的依赖, 可以使用 `openbayes_requirements.txt`。 [了解更多](#)

<input checked="" type="checkbox"/> Numpy	<input checked="" type="checkbox"/> Pandas	<input checked="" type="checkbox"/> scikit-learn	<input checked="" type="checkbox"/> Xgboost
<input checked="" type="checkbox"/> Keras	<input checked="" type="checkbox"/> OpenCV	<input checked="" type="checkbox"/> Onnx	<input checked="" type="checkbox"/> spaCy
<input checked="" type="checkbox"/> Pillow			

取消创建 审核并执行 下一步: 数据绑定

4. 在弹出的页面中点击下一步

新建训练: 基本信息 ✓ > 数据绑定 ✓ > 接入方式 ✓ > 审核

绑定数据

选择要绑定的数据仓库或输出，最多可绑定五个，[了解更多](#)

/input0

←

键入关键词进行搜索...

- 绑定到 /input 目录: 只读绑定，无法修改

- 绑定到 /output 目录: 支持读写，但拷贝过程较慢，影响容器启动速度

导入外部仓库

☐ 绑定外部 Git 仓库

绑定外部 Git 仓库（如 GitHub、Gitee）。并将外部数据导入至 /output。只支持 HTTPS 协议

取消创建

上一步

审核并执行

下一步: 接入方式

5. 点击下一步

新建训练: 基本信息 ✓ > 数据绑定 ✓ > 接入方式 ✓ > 审核

接入方式

选择容器的执行方式

工作空间
交互编辑器，推荐使用

Python 脚本
适合一次性执行场景

一种交互式的代码执行工具，创建工作空间后我们可以在其上做一些即时的命令执行和尝试性的工作。适合早期模型的调试和构建。相关的使用手册见[官方文档](#)

取消创建

上一步

审核并执行

下一步: 审核

6. 点击执行

新建训练: 基本信息 ✓ > 数据绑定 ✓ > 接入方式 ✓ > 审核

容器名称

cnnMnistDemo

容器描述

实现CNN网络识别MNIST手写数字数据集

算力

NVIDIA VGPU ①

8.59 GB 显存

2 小时 54 分钟可用

vgpu

镜像

pytorch

1.3.1

接入方式

工作空间

绑定数据

未绑定任何数据

取消创建

上一步

执行

7. 在工作空间界面中导入实验的文件 char_rnn_classification_tutorial.ipynb，data.zip

二、姓名分类

1. 任务描述:

CLASSIFYING NAMES WITH A CHARACTER-LEVEL RNN:

构建和训练一个基本的 RNN 来对单词进行分类。

将单词读取为一系列字符，在每一步输出预测和“隐藏状态”，将其先前的隐藏状态馈送到每个下一步。我们将最终预测作为输出，即单词属于哪个类。

数据集包含来自 18 种语言的几千个姓氏，rnn 模型训练完成后，根据拼写预测一个名字来自哪种语言

```
> Dovesky
(-0.20) Russian
(-2.82) Polish
(-2.96) Czech

> Jackson
(-0.78) Scottish
(-1.63) Russian
(-1.66) English

> Satoshi
(-1.22) Japanese
(-1.53) Italian
(-2.38) Polish
```

2. 流程:

- (1) 数据读入、预处理
- (2) 初始化模型、优化器
- (3) 训练、测试

3. 参考: https://pytorch.org/tutorials/intermediate/char_rnn_classification_tutorial.html
<https://www.bilibili.com/video/BV1Ui4y1t78N?p=1>

4. 代码: char_rnn_classification_tutorial.ipynb

5. 数据集: <https://download.pytorch.org/tutorial/data.zip>

三、作业安排

完成上面姓名分类任务，5月10日之前厦门大学 SPOC 平台中提交一份实验报告。