

机器学习PAI

在线预测功能介绍

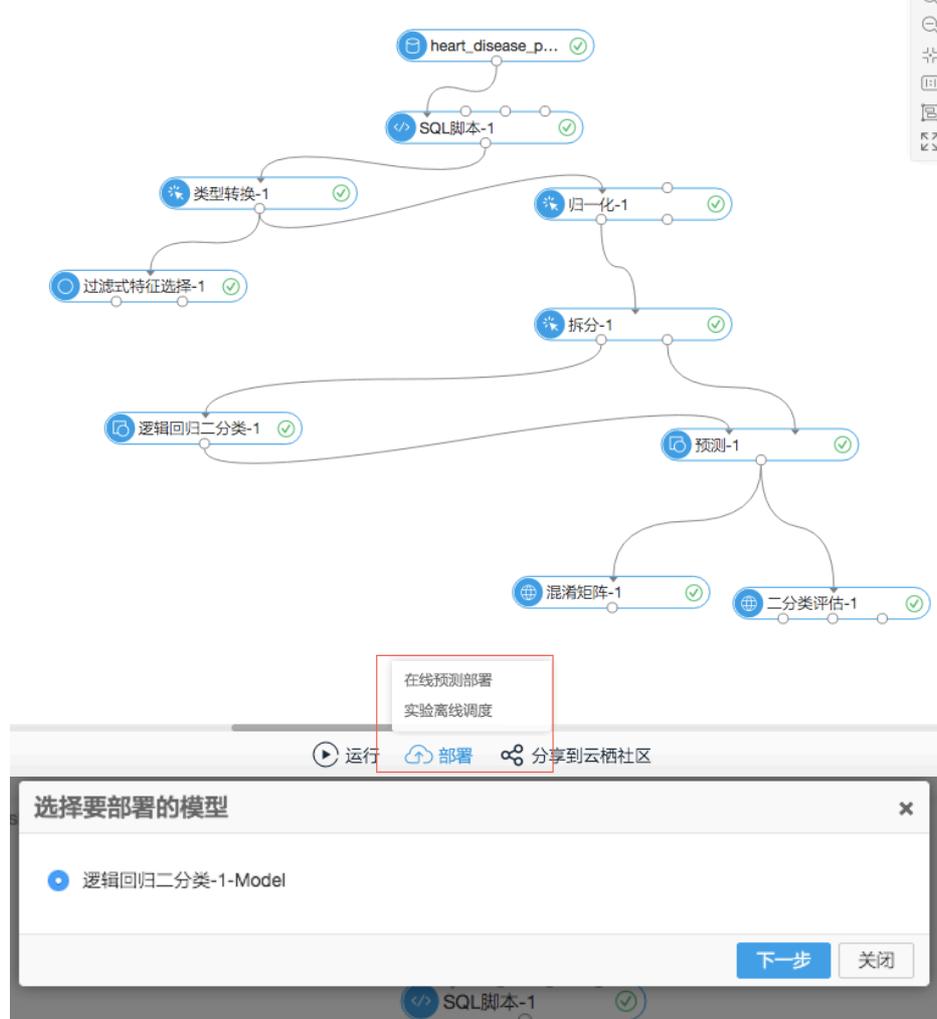
在线预测功能介绍

如何在线实时进行机器学习模型调用

机器学习PAI生成的模型可以通过在线部署方式生成API供其它业务调用。

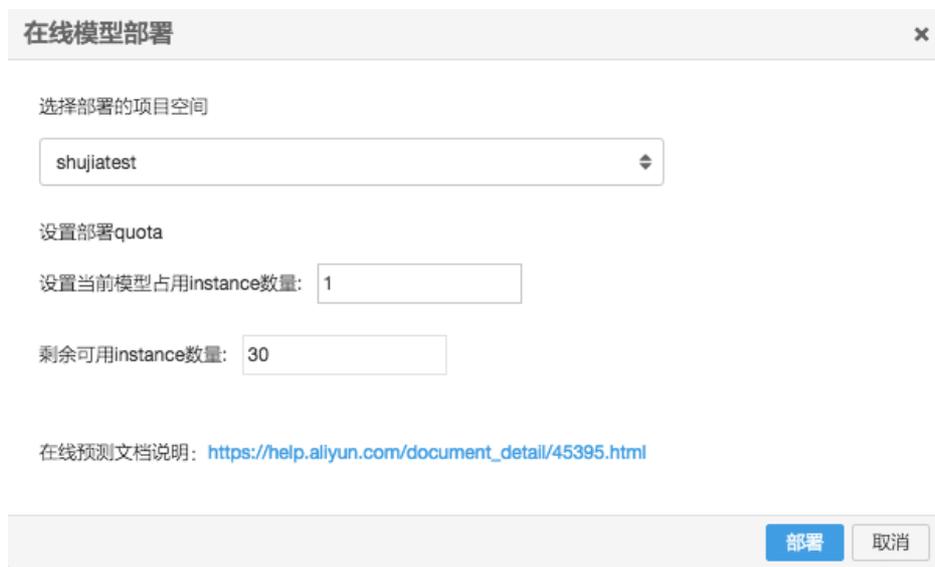
1.选择部署模型

我们以心脏病预测案例为例，具体实现可以参考https://help.aliyun.com/document_detail/34929.html，实验生成一个逻辑回归模型，是用在线预测可以在当前实验点击“部署”按钮，选择“在线预测部署”。



2.配置模型部署信息

进入模型配置页：



选择对应的项目空间，如果是第一次使用需要开通在线预测权限，权限申请是实时开通。下面详细解释instance的定义：

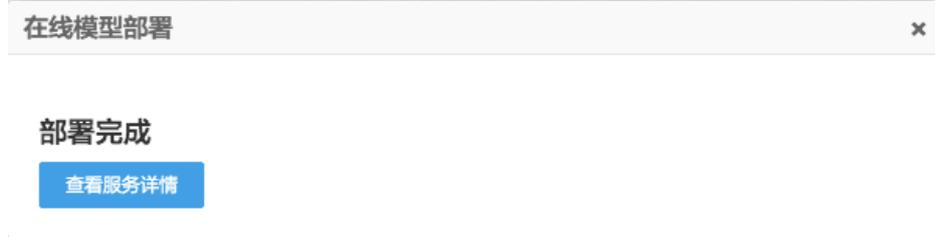
每个项目默认包含30个instance，可提工单扩容。删除已部署模型会释放当前模型的instance。

instance决定模型的QPS，每个instance为1核2G内存。

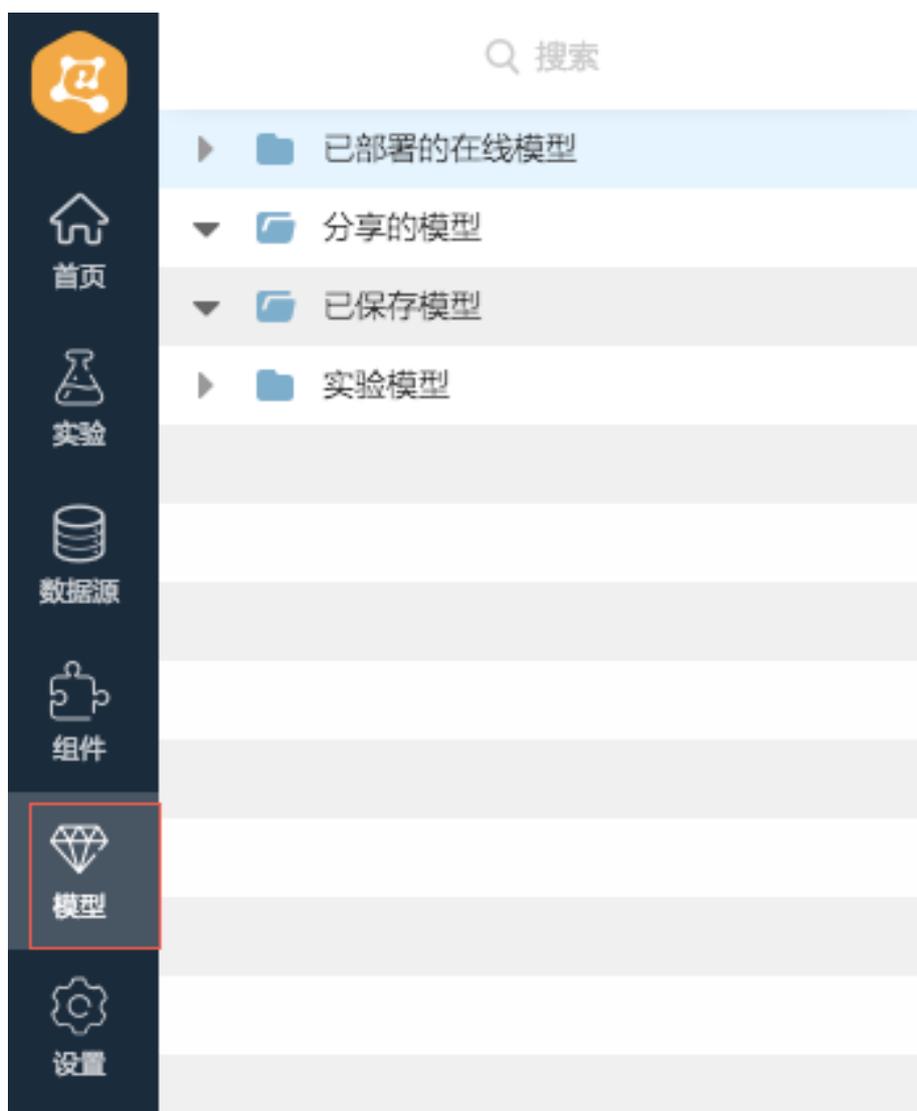
- 单个模型的instance部署限制是[1,15]。

3.模型管控

模型部署完成可以进入如下界面进行管理，新部署模型可以在“查看模型详情”进行查看。



已经部署的模型可以在“已部署在线模型”里进行管理，



模型管理界面，版本表示的是同一模型多次部署的区分，通过下图红框可以拿到模型所在的项目和模型名称：

在线预测部署

当前模型状态: 部署成功 当前版本: 0 部署时间: 2017-06-22 11:19:12 请查看页面下方信息进行

接口调用。如需更新, 请点击:

重新部署

删除当前版本

模型调试

查看历史版本信息, 请点击版本进行切换, 重新部署新增预测服务, 不会覆盖原有服务

接口模式

返回样例

帮助: https://help.aliyun.com/document_detail/45395.html

预测服务endpoint: <http://prediction.odps.aliyun.com>

部署project: shequ

在线模型名称: xlab_m_logisticregress_520728_v0

接口方式: Restful Api支持Json和Protobuf

返回格式: JSON/XML

接口样例:

POST

http://prediction.odps.aliyun.com/projects/shequ/onlinemodels/xlab_m_logisticregress_520728_v0

HTTP/1.1

Authorization: ODPS

AccessId:AccessKey

Date: Tue, 31 Mar 2015 06:32:27 GMT

Content-Type: application/json

4.模型调试

模型调试页面可以帮助用户了解在线预测请求参数的书写规范, 进入模型调试页面。

| API调试: 机器学习

您可以通过调用API来实现对您为数加订购的官方服务的调用, 这个工具帮助你快速入门, 详细请查看[机器学习API说明](#)、[数加平台API校验规则](#) (数加平台相关)。

接口名称:

请求方法:

请求地址:
请求地址可以自行加上入参, 例如<http://example.com?param1=123¶m2=456>

请求Body:

Access Key ID:
请使用团队管理员的AK, 管理员帐号可以到[成员管理](#)查看。阿里云AK可到[Access Key管理](#)查看。

Access Key Secret:

返回结果:

- 请求地址 : [https://dtplus-cn-shanghai.data.aliyuncs.com/dataplus_261422/pai/prediction/projects/\\$项目名称/onlinemodels/\\$模型名称](https://dtplus-cn-shanghai.data.aliyuncs.com/dataplus_261422/pai/prediction/projects/$项目名称/onlinemodels/$模型名称)
- 请求body为json串, 以本文逻辑回归算法为例, 需要填写每个特征的信息, 特征名字需要与模型表特

征名对应，常数列不用写。dataValue表示预测集对应特征的取值。dataType表示数值类型，dataType定义如下：

数据类型	dataType
bool	1
int32	10
int64	20
float	30
double	40
string	50

本案例body范例：

```
{
  "inputs": [
    {
      "sex": {
        "dataType": 40,
        "dataValue": 1
      },
      "cp": {
        "dataType": 40,
        "dataValue": 1
      },
      "fbs": {
        "dataType": 40,
        "dataValue": 1
      },
      "restecg": {
        "dataType": 40,
        "dataValue": 1
      },
      "exang": {
        "dataType": 40,
        "dataValue": 1
      },
      "slop": {
        "dataType": 40,
        "dataValue": 1
      },
      "thal": {
        "dataType": 40,
        "dataValue": 1
      },
      "age": {
        "dataType": 40,
        "dataValue": 1
      },
      "trestbps": {
        "dataType": 40,
        "dataValue": 1
      }
    }
  ]
}
```

```
},  
"chol": {  
  "dataType": 40,  
  "dataValue": 1  
},  
"thalach": {  
  "dataType": 40,  
  "dataValue": 1  
}  
}  
]  
}
```

可以获得返回：

```
- - - - - 请求 - - - - -  
- - - - - 返回 - - - - -  
状态码：200  
返回Body： {  
  "outputs": [  
    {  
      "outputLabel": "1",  
      "outputMulti": {  
        "0": 0.01351125016100008,  
        "1": 0.9864887498389999  
      },  
      "outputValue": {  
        "dataType": 40,  
        "dataValue": 0.9864887498389999  
      }  
    }  
  ]  
}
```

API调用方法：https://help.aliyun.com/document_detail/30245.html