

ModelArts

自动学习

文档版本 01
发布日期 2024-04-30



版权所有 © 华为云计算技术有限公司 2024。保留一切权利。

非经本公司书面许可，任何单位和个人不得擅自摘抄、复制本文档内容的部分或全部，并不得以任何形式传播。

商标声明



HUAWEI和其他华为商标均为华为技术有限公司的商标。

本文档提及的其他所有商标或注册商标，由各自的所有人拥有。

注意

您购买的产品、服务或特性等应受华为云计算技术有限公司商业合同和条款的约束，本文档中描述的全部或部分产品、服务或特性可能不在您的购买或使用范围之内。除非合同另有约定，华为云计算技术有限公司对本文档内容不做任何明示或暗示的声明或保证。

由于产品版本升级或其他原因，本文档内容会不定期进行更新。除非另有约定，本文档仅作为使用指导，本文档中的所有陈述、信息和建议不构成任何明示或暗示的担保。

目录

1 自动学习（新版）	1
1.1 自动学习简介	1
1.2 图像分类	2
1.2.1 准备数据	2
1.2.2 创建项目	4
1.2.3 数据标注	6
1.2.4 模型训练	9
1.2.5 部署上线	10
1.3 物体检测	12
1.3.1 准备数据	12
1.3.2 创建项目	14
1.3.3 数据标注	16
1.3.4 模型训练	20
1.3.5 部署上线	22
1.4 预测分析	24
1.4.1 准备数据	24
1.4.2 创建项目	26
1.4.3 模型训练	28
1.4.4 部署上线	30
1.5 声音分类	31
1.5.1 准备数据	31
1.5.2 创建项目	32
1.5.3 数据标注	34
1.5.4 模型训练	36
1.5.5 部署上线	37
1.6 文本分类	39
1.6.1 准备数据	39
1.6.2 创建项目	40
1.6.3 数据标注	42
1.6.4 模型训练	44
1.6.5 部署上线	46
1.7 使用窍门	48
1.7.1 创建项目时，如何快速创建 OBS 桶及文件夹？	48

1.7.2 自动学习生成的模型，存储在哪里？支持哪些其他操作？	49
2 自动学习（旧版）	50
2.1 自动学习简介	50
2.2 图像分类	51
2.2.1 准备数据	51
2.2.2 创建项目	52
2.2.3 数据标注	55
2.2.4 模型训练	57
2.2.5 部署上线	60
2.3 物体检测	62
2.3.1 准备数据	62
2.3.2 创建项目	65
2.3.3 数据标注	67
2.3.4 模型训练	70
2.3.5 部署上线	72
2.4 预测分析	75
2.4.1 准备数据	75
2.4.2 创建项目	77
2.4.3 选择标签列	78
2.4.4 模型训练	79
2.4.5 部署上线	81
2.5 声音分类	82
2.5.1 准备数据	83
2.5.2 创建项目	83
2.5.3 数据标注	86
2.5.4 模型训练	87
2.5.5 部署上线	89
2.6 文本分类	90
2.6.1 准备数据	90
2.6.2 创建项目	92
2.6.3 数据标注	94
2.6.4 模型训练	96
2.6.5 部署上线	98
2.7 使用窍门	100
2.7.1 创建项目时，如何快速创建 OBS 桶及文件夹？	100
2.7.2 增加的数据，如何在自动学习项目中查看？	101
2.7.3 自动学习项目中，如何进行增量训练？	102
2.7.4 自动学习生成的模型，存储在哪里？支持哪些其他操作？	103
2.7.5 将旧版项目升级到新版	104

1 自动学习（新版）

1.1 自动学习简介

自动学习功能介绍

ModelArts自动学习是帮助人们实现AI应用的低门槛、高灵活、零代码的定制化模型开发工具。自动学习功能根据标注数据自动设计模型、自动调参、自动训练、自动压缩和部署模型。开发者无需专业的开发基础和编码能力，只需上传数据，通过自动学习界面引导和简单操作即可完成模型训练和部署。

当前自动学习支持快速创建图像分类、物体检测、预测分析、声音分类和文本分类模型的定制化开发。可广泛应用在工业、零售安防等领域。

- 图像分类：识别图片中物体的类别。
- 物体检测：识别出图片中每个物体的位置和类别。
- 预测分析：对结构化数据做出分类或数值预测。
- 声音分类：对环境中不同声音进行分类识别。
- 文本分类：识别一段文本的类别。

自动学习流程介绍

使用ModelArts自动学习开发AI模型无需编写代码，您只需上传数据、创建项目、完成数据标注、发布训练、然后将训练的模型部署上线。具体流程请参见图1-1。新版自动学习中，该流程可完全由Workflow进行承载，如图1-2。开发者可以通过Workflow进行有向无环图（Directed Acyclic Graph, DAG）的开发，整个DAG的执行就是有序的任务执行模板，依次执行从数据标注、数据集版本发布、模型训练、模型注册到服务部署环节。若想了解更多关于Workflow您可以参考[Workflow简介](#)。

图 1-1 自动学习操作流程

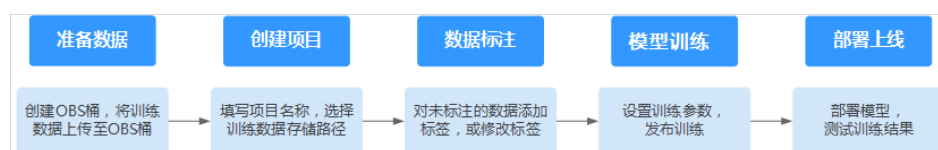


图 1-2 Workflow 运行流程



项目类型介绍

- **图像分类**

图像分类项目，是对图像进行分类。需要添加图片并对图像进行分类标注，完成图片标注后开始模型训练，即可快速生成图像分类模型。可应用于商品的自动分类、运输车辆种类识别和残次品的自动分类等。例如质量检查的场景，则可以上传产品图片，将图片标注“合格”、“不合格”，通过训练部署模型，实现产品的质检。

- **物体检测**

物体检测项目，是检测图片中物体的类别与位置。需要添加图片，用合适的框标注物体作为训练集，进行训练输出模型。适用于一张图片中要识别多个物体或者物体的计数等。可应用于园区人员穿戴规范检测和物品摆放的无人巡检。

- **预测分析**

预测分析项目，是一种针对结构化数据的模型自动训练应用，能够对结构化数据进行分类或者数据预测。可用于用户画像分析，实现精准营销。也可应用于制造设备预测性维护，根据设备实时数据的分析，进行故障识别。

- **声音分类**

声音分类项目，是识别一段音频中是否包含某种声音。可应用于生产或安防场景的异常声音监控。

- **文本分类**

文本分类项目，识别一段文本的类别。可应用于情感分析或新闻分类等场景。

1.2 图像分类

1.2.1 准备数据

使用ModelArts自动学习构建模型时，您需要将数据上传至对象存储服务（OBS）中。OBS桶需要与ModelArts在同一区域。

数据集要求

- 保证图片质量：不能有损坏的图片，目前支持的格式包括jpg、jpeg、bmp、png。
- 不要把明显不同的多个任务数据放在同一个数据集内。
- 每一类数据尽量多，尽量均衡。期望获得良好效果，图像分类项目中，至少有两种以上的分类，每种分类的样本不少于20张。
- 为了保证模型的预测准确度，训练样本跟真实使用场景尽量相似。

- 为保证模型的泛化能力，数据集尽量覆盖可能出现的各种场景。

数据上传至 OBS

在本文档中，采用通过OBS管理控制台将数据上传至OBS桶。

上传OBS的文件规范：

- 文件名规范：不能有+、空格、制表符。
- 如不需要提前上传训练数据，请创建一个空文件夹用于存放工程后期生成的文件。如：“/bucketName/data-cat”。
- 如需要提前上传待标注的图片，请创建一个空文件夹，然后将图片文件保存在该文件夹下，图片的目录结构如：“/bucketName/data-cat/cat.jpg”。
- 如您将已标注好的图片上传至OBS桶，请按照如下规范上传。
 - 图像分类数据集要求将标注对象和标注文件存储在同一目录，并且一一对应，例如标注对象文件名为“10.jpg”，那么标注文件的文件名应为“10.txt”。

数据文件存储示例：

```
<dataset-import-path>
├── 10.jpg
├── 10.txt
├── 11.jpg
├── 11.txt
├── 12.jpg
└── 12.txt
```

- 只支持JPG、JPEG、PNG、BMP格式的图片。在OBS管理控制台上传时，单张图片的大小不能超过5MB，单次上传的图片总大小不能超过8MB，数据量大时推荐使用OBS Browser+上传。
- 标签名是由中文、大小写字母、数字、中划线或下划线组成，且不超过32位的字符串。
- 图像分类标签“.txt”规范如下。

一行一个标签：

```
flower
book
...
```

上传OBS操作步骤：

执行如下操作，将数据上传到OBS中，以便用于模型训练和构建。

1. 登录OBS管理控制台，在ModelArts同一区域内[创建桶](#)。如果已存在可用的桶，需确保OBS桶与ModelArts在同一区域。
2. 参考[上传文件](#)，将本地数据上传至OBS桶中。如果您的数据较多，推荐OBS Browser+上传数据或上传文件夹。上传的数据需满足此类型自动学习项目的数据集要求。

📖 说明

在上传数据时，请选择非加密桶进行上传，否则会由于加密桶无法解密导致后期的训练失败。

创建数据集

数据准备完成后，需要创建相应项目支持的类型的数据集，具体操作请参考[创建数据集](#)。

1.2.2 创建项目

ModelArts自动学习，包括图像分类、物体检测、预测分析、声音分类和文本分类项目。您可以根据业务需求选择创建合适的项目。您需要执行如下操作来创建自动学习项目。

创建项目

1. 登录ModelArts管理控制台，在左侧导航栏单击“自动学习”，进入新版自动学习页面。
2. 在您需要的自动学习项目列表中，单击“创建项目”，进入创建自动学习项目界面。
3. 在创建自动学习项目页面，参考表1-1填写相应参数。

表 1-1 参数说明

参数	说明
“名称”	项目的名称。 <ul style="list-style-type: none">● 名称只能包含数字、字母、下划线和中划线，长度不能超过64位且不能为空。● 名称请以字母开头。● 名称不允许重复。
“描述”	对项目的简要描述。
“数据集”	可在右侧下拉框选择已有数据集，或单击“创建数据集”前往新建数据集。 <ul style="list-style-type: none">● 已有数据集：在“数据集”右侧的下拉框中选择，仅展示同类型的数据集供选择。● 创建数据集：前往创建数据集页面创建一个新的数据集。具体可参考如何创建数据集。
“输出路径”	选择自动学习数据输出的统一OBS路径。 说明 “输出路径”是存储自动学习在运行过程中所有产物的路径。
“训练规格”	选择自动学习训练节点所使用的资源规格，以实际界面显示为准，将会根据不同的规格计费。 说明 <ul style="list-style-type: none">● 只有北京四区域支持限时免费规格。● 若您购买了套餐包，可优先选择您对应规格的套餐包，在“配置费用”处会显示您的套餐余量，以及超出的部分如何计费，请您关注，避免造成不必要的资源浪费。

4. 单击“创建项目”，图像分类项目创建成功后页面自动跳转到“自动学习 workflow”。
5. 图像分类项目的工作流，将依次运行如下节点：
 - a. 数据标注：对您的数据标注情况进行确认。

- b. 数据集版本发布：将已完成标注的数据进行版本发布。
- c. 数据校验：对您的数据集的数据进行校验，是否存在数据异常。
- d. 图像分类：将发布好的数据集版本进行训练，生成对应的模型。
- e. 模型注册：将训练后的结果注册到模型管理中。
- f. 服务部署：将生成的模型部署为在线服务。

快速查找创建好的项目

在自动学习总览页，您可以通过搜索框，根据自动学习的属性类型（项目名称）快速搜索过滤到相应的工作流，可节省您的时间。

1. 登录ModelArts管理控制台，在左侧导航栏选择自动学习，进入自动学习总览页面。
2. 在自动学习列表上方的搜索框中，根据您需要的属性类型，例如，名称、状态、项目类型、当前节点、标签等，过滤出相应的工作流。



3. 单击搜索框右侧的  按钮，可选择自动学习的基础设置，需要的显示列。
表格内容折行：默认为关闭状态，启用此能力可让表格内容自动折行，禁用此功能可截断文本。
操作列：默认为关闭状态，启用此能力可让操作列固定在最后一列永久可见。
自定义显示列：默认所有显示项全部勾选，您可以根据实际需要定义您的显示列。

图 1-3 表格显示设置



4. 单击“确定”即可按照设置好的显示列进行显示。
5. 同时可支持对自动学习项目显示页进行排序，单击表头中的箭头 ，就可对该列进行排序。

1.2.3 数据标注

由于模型训练过程需要大量有标签的图片数据，因此在模型训练之前需对没有标签的图片添加标签。通过ModelArts您可对图片进行一键式批量添加标签，快速完成对图片的标注操作，也可以对已标注图片修改或删除标签进行重新标注。

📖 说明

请确保数据集中已标注的图片不低于100张，否则会导致数据集校验环节不通过，影响您的模型训练。

项目创建完成后，将会自动跳转至新版自动学习页面，并开始运行。单击“数据标注”节点，当状态变为“等待操作”时，需要手动进行确认数据集中的数据标注情况，也可以对数据集中的数据进行标签的修改，数据的增加或删除。

图 1-4 数据标注节点状态



图片标注

1. 在新版自动学习页面的数据标注节点单击“实例详情”按钮，前往数据标注页面。

图 1-5 单击实例详情



2. 依次勾选待标注的图片，或勾选“选择当前页”选中该页面所有图片，在页面右侧进行图片标注。
3. 选中图片后，在页面右侧“添加标签”，输入“标签名”或从下拉列表中选择已添加的标签。单击“确定”，完成选中图片的标注操作。例如，您可以选择多张

图片，按照花朵种类将图片标注为“tulips”。同样选择其他未标注分类图片，将其标注为“sunflowers”、“roses”等。标注完成后，图片将存储至“已标注”页签下。

- a. 图片标注支持多标签，即一张图片可添加多个标签。
 - b. 标签名是由中文、大小写字母、数字、中划线或下划线组成。
4. 当图片目录中所有图片都完成标注后，您可以在“已标注”页签下查看已完成标注的图片，或者通过右侧的“全部标签”列表，了解当前已完成的标签名称和标签数量。

同步或添加图片

在“数据标注”节点单击“实例详情”进入数据标注页面，数据标注的图片来源有两种，通过本地添加图片和同步OBS中的图片数据。

图 1-6 添加本地图片



图 1-7 同步 OBS 图片数据



- **添加数据**：您可以将本地图片快速添加到ModelArts，同时自动上传至创建项目时所选择的OBS路径中。单击“添加数据”，根据弹出的对话框的引导，输入正确的数据并添加。
- **同步新数据**：将图片数据上传至创建项目时指定的OBS目录，然后单击“同步新数据”，快速将原OBS目录中的新数据添加到ModelArts数据集。
- **删除图片**：您可以依次单击选中图片进行删除，也可以勾选“选择当前页”对该页面所有图片进行删除。

📖 说明

所有的删除操作均不可恢复，请谨慎操作。

修改标注

当数据完成标注后，您还可以进入已标注页签，对已标注的数据进行修改。

- **基于图片修改**

在数据标注页面，单击“已标注”页签，然后在图片列表中选中待修改的图片（选择一个或多个）。在右侧标签信息区域中对图片信息进行修改。



- 添加标签：在“标签名”右侧文本框中，选择已有标签或输入新的标签名，然后单击 ，为选中图片增加标签。
- 修改标签：在“选中文件标签”区域中，单击操作列的编辑图标，然后在文本框中输入正确的标签名，然后单击确定图标完成修改。

图 1-8 编辑标签

选中文件标签

标签名称	标签数量	操作
可回收	1	 

- 删除标签：在“选中文件标签”区域中，单击操作列的  删除该标签。

- **基于标签修改**

在数据标注概览页，单击右侧的“标签管理”，即可显示全部标签的信息。

图 1-9 全部标签的信息

<input type="checkbox"/> 添加标签	<input type="checkbox"/> 删除标签		
标签名称	属性	标签颜色	操作
<input type="checkbox"/> 可回收	矩形框		 

- 修改标签：在需要修改的标签的“操作”列，单击“修改”，输入修改后的标签，单击“确定”即可。
- 删除标签：选择对应的标签，单击操作列的“删除”，在弹出的“删除标签”对话框中单击“确定”即可删除对应的标签。

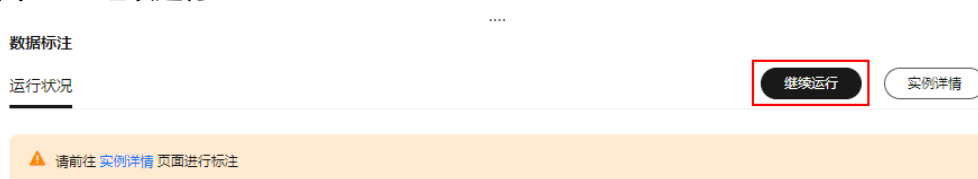
说明

删除后无法再恢复，请谨慎操作。

继续运行

完成数据的确认之后，返回自动学习的页面，在数据标注节点单击“继续运行”， workflow 将会继续依次运行直到所有节点运行成功。

图 1-10 继续运行



1.2.4 模型训练

完成图片标注后，可进行模型的训练。模型训练的目的在于得到满足需求的图像分类模型。请参考[前提条件](#)确保已标注的图片符合要求，否则数据集校验将会不通过。

前提条件

1. 请确保您的数据集中的已标注的图片不低于100张。
2. 请确保您的数据集中至少存在2种以上的图片分类，且每种分类的图片不少于5张。

操作步骤


1. 参考[数据标注](#)章节，确保您的数据已全部完成标注。
2. 在新版自动学习页面，单击数据标注节点的“继续运行”按钮，然后等待 workflow 按顺序进入训练节点即可。
3. 模型将会自动进入训练，无需人工介入，训练时间相对较长，建议您耐心等待。如果关闭或退出此页面，系统仍然在执行训练操作。
4. 在“图像分类”节点中，待训练状态由“运行中”变为“运行成功”，即完成了模型的自动训练。
5. 训练完成后，您可以单击“图像分类”节点上方的按钮，查看相关指标信息，如“准确率”、“评估结果”等。评估结果参数说明请参见[表1-2](#)。

图 1-11 模型评估报告



Metric	Value
accuracy	0.93
balanced_accuracy	0.77
f1	0.5
mcc	0.509
precision	1
recall	0.33
roc_auc	0.952

表 1-2 评估结果参数说明

参数名称	参数含义	说明
recall	召回率	被用户标注为某个分类的所有样本中，模型正确预测为该分类的样本比率，反映模型对正样本的识别能力。
precision	精确率	被模型预测为某个分类的所有样本中，模型正确预测的样本比率，反映模型对负样本的区分能力。

参数名称	参数含义	说明
accuracy	准确率	所有样本中，模型正确预测的样本比率，反映模型对样本整体的识别能力。
f1	F1值	F1值是模型精确率和召回率的加权调和平均，用于评价模型的好坏，当F1较高时说明模型效果较好。

📖 说明

同一个自动学习项目可以训练多次，每次训练会注册一个新的AI应用版本。如第一次训练版本号“0.0.1”，下一个版本为“0.0.2”。基于训练版本可以对训练模型进行管理。当训练的模型达到目标后，再执行部署上线的操作。

1.2.5 部署上线

部署上线

部署上线操作即将模型部署为在线服务，并且提供在线的测试UI与监控能力。完成模型训练后，可选择准确率理想且训练状态为“运行成功”的版本部署上线。具体操作步骤如下。

1. 在“运行节点”页面中，待服务部署节点的状态变为“等待输入”时，双击“服务部署”进入配置详情页，完成资源的参数配置操作。
2. 在服务部署页面，选择部署上线使用的资源规格。
 - AI应用来源：默认为生成的AI应用。
 - 选择AI应用及版本：自动匹配当前使用的AI应用版本，支持选择版本。
 - 资源池：默认公共资源池。
 - 分流：默认为100，输入值必须是0-100之间。
 - 计算节点规格：请根据界面显示的列表，选择可用的规格，置灰的规格表示当前环境无法使用。如果公共资源池下规格为空数据，表示当前环境无公共资源。建议使用专属资源池，或者联系系统管理员创建公共资源池。
 - 计算节点个数：默认为1，输入值必须是1-5之间的整数。
 - 是否自动停止：启用该参数并设置时间后，服务将在指定时间后自动停止。如果不启用此参数，在线服务将一直运行，同时一直收费，自动停止功能可以帮您避免产生不必要的费用。默认开启自动停止功能，且默认值为“1小时后”。

目前支持设置为“1小时后”、“2小时后”、“4小时后”、“6小时后”、“自定义”。如果选择“自定义”的模式，可在右侧输入框中输入1~24范围内的任意整数。

📖 说明

若您购买了套餐包，计算节点规格可选择您的套餐包，同时在“配置费用”页签还可查看您的套餐包余量以及超出部分的计费方式，请您务必关注，避免造成不必要的资源浪费。

3. 完成资源配置后，单击“继续运行”，服务部署节点将继续运行，直至状态变为“运行成功”，至此，已将AI应用部署为在线服务。

服务测试

服务部署节点运行成功后，单击“实例详情”可跳转至对应的在线服务详情页面。单击“预测”页签，进行服务测试。您可以通过调用代码对服务进行测试，根据部署服务类型的不同，具体操作详情参见[访问在线服务](#)。

图 1-12 服务测试



下面的测试，是您在自动学习图像分类项目页面将模型部署上线之后进行服务测试的操作步骤。

1. 模型部署完成后，“在服务部署”节点，单击“实例详情”按钮，进入服务预测界面，在“预测”页签单击“上传”，选择本地图片进行测试。
2. 单击“预测”进行测试，预测完成后，右侧“预测结果”区域输出标签名称“sunflowers”和检测的评分。如模型准确率不满足预期，可在“数据标注”页签中添加图片并进行标注，重新进行模型训练及部署上线。预测结果中的参数说明请参见[表1-3](#)。如果您对模型预测结果满意，可根据界面提示调用接口访问在线服务，操作指导请参见“[访问在线服务](#)”。

目前只支持jpg、jpeg、bmp、png格式的图片。

图 1-13 预测结果



表 1-3 预测结果中的参数说明

参数	说明
predicted_label	表示图片预测的标签。
scores	表示Top5标签的预测置信度。

📖 说明

- 由于“运行中”的在线服务将持续耗费资源，如果不需再使用此在线服务，建议在“在线服务”的操作列单击“更多>停止”，避免产生不必要的费用。如果需要继续使用此服务，可单击“启动”恢复。
- 如果您启用了自动停止功能，服务将在指定时间后自动停止，不再产生费用。

1.3 物体检测

1.3.1 准备数据

使用ModelArts自动学习构建模型时，您需要将数据上传至对象存储服务（OBS）中。OBS桶需要与ModelArts在同一区域。

数据集要求

- 保证图片质量：不能有损坏的图片；目前支持的格式包括jpg、jpeg、bmp、png。
- 不要把明显不同的多个任务数据放在同一个数据集内。
- 为了保证模型的预测准确度，训练样本跟真实使用场景尽量相似。
- 为保证模型的泛化能力，数据集尽量覆盖可能出现的各种场景。
- 物体检测数据集中，如果标注框坐标超过图片，将无法识别该图片为已标注图片。

数据上传至 OBS

在本文档中，采用通过OBS管理控制台将数据上传至OBS桶。

上传OBS的文件规范：

- 文件名规范，不能有中文，不能有+、空格、制表符。
- 如不需要提前上传训练数据，请创建一个空文件夹用于存放工程后期生成的文件。如：“/bucketName/data-cat”。
- 如需要提前上传待标注的图片，请创建一个空文件夹，然后将图片文件保存在该文件夹下，图片的目录结构如：“/bucketName/data-cat/cat.jpg”。
- 如您将已标注好的图片上传至OBS桶，请按照如下规范上传。
 - 物体检测数据集要求用户将标注对象和标注文件存储在同一目录，并且一一对应。例如标注对象文件名为“IMG_20180919_114745.jpg”，那么标注文件的文件名应为“IMG_20180919_114745.xml”。

物体检测的标注文件需要满足PASCAL VOC格式，格式详细说明请参见 [表 1-4](#)。

数据存储示例：

```
<dataset-import-path>
|
|  IMG_20180919_114732.jpg
|  IMG_20180919_114732.xml
|  IMG_20180919_114745.jpg
|  IMG_20180919_114745.xml
|  IMG_20180919_114945.jpg
|  IMG_20180919_114945.xml
```


- 只支持JPG、JPEG、PNG、BMP格式的图片，在OBS管理控制台上传时，单张图片的大小不能超过5MB，单次上传的图片总大小不能超过8MB，数据量大时推荐使用OBS Browser+上传。
- 标签名是由中文、大小写字母、数字、中划线或下划线组成，且不超过32位的字符串。

表 1-4 PASCAL VOC 格式说明

字段	是否必选	说明
folder	是	表示数据源所在目录。
filename	是	被标注文件的文件名。
size	是	表示图像的像素信息。 <ul style="list-style-type: none"> • width: 必选字段，图片的宽度。 • height: 必选字段，图片的高度。 • depth: 必选字段，图片的通道数。
segmented	是	表示是否用于分割。
object	是	表示物体检测信息，多个物体标注会有多个object体。 <ul style="list-style-type: none"> • name: 必选字段，标注内容的类别。 • pose: 必选字段，标注内容的拍摄角度。 • truncated: 必选字段，标注内容是否被截断（0表示完整）。 • occluded: 必选字段，标注内容是否被遮挡（0表示未遮挡）。 • difficult: 必选字段，标注目标是否难以识别（0表示容易识别）。 • confidence: 可选字段，标注目标的置信度，取值范围0-1之间。 • bndbox: 必选字段，标注框的类型，标注信息请参见 表1-5。

表 1-5 标注框类型描述

type	形状	标注信息
bndbox	矩形框	左上和右下两个点坐标。 <xmin>100<xmin> <ymin>100<ymin> <xmax>200<xmax> <ymin>200<ymin>

标注文件示例：

```
<annotation>
  <folder>test_data</folder>
  <filename>260730932.jpg</filename>
  <size>
    <width>767</width>
    <height>959</height>
    <depth>3</depth>
  </size>
  <segmented>0</segmented>
  <object>
    <name>bag</name>
    <pose>Unspecified</pose>
    <truncated>0</truncated>
    <occluded>0</occluded>
    <difficult>0</difficult>
    <bndbox>
      <xmin>108</xmin>
      <ymin>101</ymin>
      <xmax>251</xmax>
      <ymax>238</ymax>
    </bndbox>
  </object>
</annotation>
```

上传OBS的操作步骤：

执行如下操作，将数据导入到您的数据集中，以便用于模型训练和构建。

1. 登录OBS管理控制台，在ModelArts同一区域内**创建桶**。如果已存在可用的桶，需确保OBS桶与ModelArts在同一区域。
2. 参考**上传文件**，将本地数据上传至OBS桶中。如果您的数据较多，推荐OBS Browser+上传数据或上传文件夹。上传的数据需满足此类型自动学习项目的数据集要求。

📖 说明

- 在上传数据时，请选择非加密桶进行上传，否则会由于加密桶无法解密导致后期的训练失败。
- 用于训练的图片，至少有1种以上的分类，每种分类的图片数不少50张。

创建数据集

数据准备完成后，需要创建相应项目支持的类型的数据集，具体操作请参考**创建数据集**。

1.3.2 创建项目

ModelArts自动学习，包括图像分类、物体检测、预测分析、声音分类和文本分类项目。您可以根据业务需求选择创建合适的项目。您需要执行如下操作来创建自动学习项目。

创建项目

1. 登录ModelArts管理控制台，在左侧导航栏单击“自动学习”，进入新版自动学习页面。
2. 在您需要的自动学习项目列表中，单击“创建项目”，进入创建自动学习项目界面。
3. 在创建自动学习项目页面，参考**表1-6**填写相应参数。

表 1-6 参数说明

参数	说明
“名称”	项目的名称。 <ul style="list-style-type: none"> 名称只能包含数字、字母、下划线和中划线，长度不能超过64位且不能为空。 名称请以字母开头。 名称不允许重复。
“描述”	对项目的简要描述。
“数据集”	可在右侧下拉框选择已有数据集，或单击“创建数据集”前往新建数据集。 <ul style="list-style-type: none"> 已有数据集：在“数据集”右侧的下拉框中选择，仅展示同类型的数据集供选择。 创建数据集：前往创建数据集页面创建一个新的数据集。具体可参考如何创建数据集。
“输出路径”	选择自动学习数据输出的统一OBS路径。 说明 “输出路径”是存储自动学习在运行过程中所有产物的路径。
“训练规格”	选择自动学习训练节点所使用的资源规格，以实际界面显示为准，将会根据不同的规格计费。 说明 <ul style="list-style-type: none"> 只有北京四区域支持限时免费规格。 若您购买了套餐包，可优先选择您对应规格的套餐包，在“配置费用”处会显示您的套餐余量，以及超出的部分如何计费，请您关注，避免造成不必要的资源浪费。

4. 单击“创建项目”，物体检测项目创建成功后页面自动跳转到“自动学习 workflow”。
5. 物体检测项目的工作流，将依次运行如下节点：
 - a. 数据标注：对您的数据进行标注情况确认。
 - b. 数据集版本发布：将已完成标注的数据进行版本发布。
 - c. 数据校验：对您的数据集的数据进行校验，是否存在数据异常。
 - d. 物体检测：将发布好的数据集版本进行训练，生成对应的模型。
 - e. 模型注册：将训练后的结果注册到模型管理中。
 - f. 服务部署：将生成的模型部署为在线服务。

快速查找创建好的项目

在自动学习总览页，您可以通过搜索框，根据自动学习的属性类型（项目名称）快速搜索过滤到相应的工作流，可节省您的时间。

1. 登录ModelArts管理控制台，在左侧导航栏选择自动学习，进入自动学习总览页面。
2. 在自动学习列表上方的搜索框中，根据您需要的属性类型，例如，名称、状态、项目类型、当前节点、标签等，过滤出相应的工作流。



- 单击搜索框右侧的  按钮，可选择自动学习的基础设置，需要的显示列。
表格内容折行：默认为关闭状态，启用此能力可让表格内容自动折行，禁用此功能可截断文本。
操作列：默认为关闭状态，启用此能力可让操作列固定在最后一列永久可见。
自定义显示列：默认所有显示项全部勾选，您可以根据实际需要定义您的显示列。

图 1-14 表格显示设置



- 单击“确定”即可按照设置好的显示列进行显示。
- 同时可支持对自动学习项目显示页进行排序，单击表头中的箭头 ，就可对该列进行排序。

1.3.3 数据标注

物体检测之前，首先需考虑如何设计标签，标签设计需要对应所检测图片的明显特征，并且选择的标签比较容易识别（画面主体物与背景区分度较高），每个标签就是对所检测图片期望识别的全部结果。物体的标签设计完成之后，基于设计好的标签准备该图片的数据，每种需识别出的标签，建议应在所有图片个数相加超过100张，若某些图片的标签具有相似性，则需要更多的图片。用于训练的图片，至少有1种以上的分类，每种分类的图片数不少50张。


- 标注时，类内方差尽量要小。即相同类别的标注，尽量近似；不同类别的标注，尽量保持差距较大。
- 标记的每个标签尽量和背景有较大的区分度。
- 物体检测标注，需要保证目标框内物体的完整性；针对图片中存在多个物体的情形，做到不重标、不漏标。

项目创建完成后，将会自动跳转至新版自动学习页面，并开始运行，当数据标注节点的状态变为“等待操作”时，需要手动进行确认数据集中的数据标注情况，也可以对数据集中的数据进行标签的修改，数据的增加或删除。

图 1-15 数据标注节点状态

数据标注

运行状况

 请前往 [实例详情](#) 页面进行标注

属性

状态 ● 等待操作

图片标注

1. 在新版自动学习页面单击“实例详情”按钮，前往数据标注页面。单击任意一张图片，进入图片标注界面。
2. 用鼠标框选图片中的物体所在区域，然后在弹出的对话框中选择标签颜色，输入标签名称，例如此示例中的“yunbao”，按“Enter”键完成此标签的添加。标注完成后，左侧图片目录中此图片的状态将显示为“已标注”。

数据标注的更多说明：

- 您可以在图片上方或下方单击左右切换键，或者按键盘的左右方向键，选择其他图片，重复上述操作继续进行图片标注。如果一张图片有多个物体，您可以标注多处。
- 同一个物体检测自动学习项目内，可以增加多个标签，且标签可选择不同颜色，方便识别。使用鼠标完成物体框选后，在弹出的对话框中，选择新的颜色，输入新的标签名称，即可添加一个新的标签。
- 自动学习项目中，物体检测仅支持矩形标注框。在“数据管理”功能中，物体检测类型的数据集，支持更多类型的标注框。
- 在标注窗口中，您可以滚动鼠标，放大或缩小图片，方便您快速定位到物体位置。

图 1-16 物体检测图片标注



说明

“物体检测”类型的数据集，在标注时，支持在一张图片中添加多个标注框以及标签。需要注意的是，标注框不能超过图片边缘。

- 当图片目录中所有图片都完成标注后，返回“自动学习 workflow”页面，单击“继续运行”按钮， workflow 将会自动发布数据标注版本，并进行下一步训练步骤。

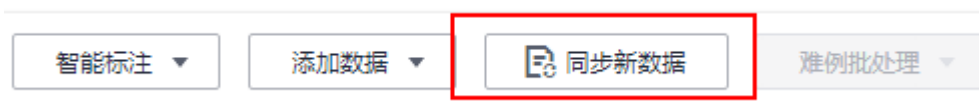
同步或添加图片

在“数据标注”节点单击“实例详情”进入数据标注页面，数据标注的图片来源有两种，通过本地添加图片和同步 OBS 中的图片数据。

图 1-17 添加本地图片



图 1-18 同步 OBS 图片数据



- **添加数据**：您可以将本地图片快速添加到ModelArts，同时自动上传至创建项目时所选择的OBS路径中。单击“添加数据”，根据弹出的对话框的引导，输入正确的数据并添加。
- **同步新数据**：将图片数据上传至创建项目时指定的OBS目录，然后单击“同步新数据”，快速将原OBS目录中的新数据添加到ModelArts数据集。
- **删除图片**：您可以依次单击选中图片进行删除，也可以勾选“选择当前页”对该页面所有图片进行删除。

📖 说明

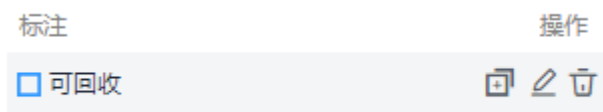
所有的删除操作均不可恢复，请谨慎操作。

修改标注

当数据完成标注后，您还可以进入已标注页签，对已标注的数据进行修改。

- **基于图片修改**
在数据集详情页面，单击“已标注”页签，然后在图片列表中选中待修改的图片，在右侧“标注”区域中对图片信息进行修改。
 - **修改标签**：“标注”区域中，单击编辑按钮，在文本框中输入正确的标签名，然后单击确定按钮完成修改。标签颜色不支持修改。
 - **删除标签**：在“标注”区域中，单击删除按钮，即可删除此图片中的标签。标签删除后，单击页面左上角的项目名称离开标注页面。该图片会重新回到“未标注”页签。

图 1-19 编辑物体检测标签



- **基于标签修改**
在数据标注作业概览页，单击右侧的“标签管理”，进入标签管理页面，标签管理页展示所有标签信息。

图 1-20 标签管理页



- **修改标签**：单击操作列的“修改”按钮，然后在弹出的对话框中输入修改后的标签名，然后单击“确定”完成修改。修改后，之前添加了此标签的图片，都将被标注为新的标签名称。

- 删除标签：单击操作列的“删除”按钮，在弹出的对话框中，根据界面提示选择删除对象，然后单击“确定”。

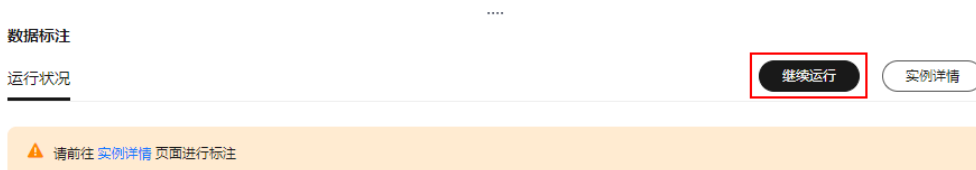
📖 说明

删除后的标签无法再恢复，请谨慎操作。

继续运行

完成数据的确认之后，返回新版自动学习的页面，在数据标注节点单击“继续运行”， workflows 将会继续依次运行直到所有节点运行成功。

图 1-21 继续运行



1.3.4 模型训练

自动学习物体检测项目，在图片标注完成后，通过模型训练得到合适的模型版本。

操作步骤

1. 在新版自动学习页面，单击项目名称进入运行总览页面，单击“数据标注”节点的“实例详情”进入数据标注页面，完成数据标注。

图 1-22 完成数据标注



2. 返回新版自动学习页面，单击数据标注节点的“继续运行”，然后等待 workflow 按顺序进入训练节点。
3. 模型将会自动进入训练，无需人工介入，训练时间相对较长，建议您耐心等待。如果关闭或退出此页面，系统仍然在执行训练操作。
4. 在“物体检测”节点中，待训练状态由“运行中”变为“运行成功”，即完成模型的自动训练。

图 1-23 运行成功




5. 训练完成后，您可以单击物体检测节点上方的  按钮，查看相关指标信息，如“准确率”、“评估结果”等。评估结果参数说明请参见表1-7。

图 1-24 模型评估报告

Metric	Value
accuracy	0.93
balanced_accuracy	0.77
f1	0.5
mcc	0.509
precision	1
recall	0.33
roc_auc	0.952

表 1-7 评估结果参数说明

参数	说明
recall: 召回率	被用户标注为某个分类的所有样本中，模型正确预测为该分类的样本比率，反映模型对正样本的识别能力。
precision: 精确率	被模型预测为某个分类的所有样本中，模型正确预测的样本比率，反映模型对负样本的区分能力。
accuracy: 准确率	所有样本中，模型正确预测的样本比率，反映模型对样本整体的识别能力。
f1: F1值	F1值是模型精确率和召回率的加权调和平均，用于评价模型的好坏，当F1较高时说明模型效果较好。

📖 说明

同一个自动学习项目可以训练多次，每次训练会注册一个新的AI应用一个版本。如第一次训练版本号“0.0.1”，下一个版本为“0.0.2”。基于训练版本可以对训练模型进行管理。当训练的模型达到目标后，再执行部署上线的操作。

1.3.5 部署上线

部署上线

部署上线操作即将模型部署为在线服务，并且提供在线的测试UI与监控能力。完成模型训练后，可选择准确率理想且训练状态为“运行成功”的版本部署上线。具体操作步骤如下。

1. 在“运行节点”页面中，待服务部署节点的状态变为“等待输入”时，双击“服务部署”进入配置详情页，完成资源的参数配置操作。
2. 在服务部署页面，选择部署上线使用的资源规格。
 - AI应用来源：默认为生成的AI应用。
 - 选择AI应用及版本：自动匹配当前使用的AI应用版本，支持选择版本。
 - 资源池：默认公共资源池。
 - 分流：默认为100，输入值必须是0-100之间。
 - 计算节点规格：请根据界面显示的列表，选择可用的规格，置灰的规格表示当前环境无法使用。如果公共资源池下规格为空数据，表示当前环境无公共资源。建议使用专属资源池，或者联系系统管理员创建公共资源池。
 - 计算节点个数：默认为1，输入值必须是1-5之间的整数。
 - 是否自动停止：启用该参数并设置时间后，服务将在指定时间后自动停止。如果不启用此参数，在线服务将一直运行，同时一直收费，自动停止功能可以帮您避免产生不必要的费用。默认开启自动停止功能，且默认值为“1小时后”。
目前支持设置为“1小时后”、“2小时后”、“4小时后”、“6小时后”、“自定义”。如果选择“自定义”的模式，可在右侧输入框中输入1~24范围内的任意整数。

📖 说明

若您购买了套餐包，计算节点规格可选择您的套餐包，同时在“配置费用”页签还可查看您的套餐包余量以及超出部分的计费方式，请您务必关注，避免造成不必要的资源浪费。

3. 完成资源配置后，单击“继续运行”，服务部署节点将继续运行，直至状态变为“运行成功”，至此，已将AI应用部署为在线服务。

服务测试

服务部署节点运行成功后，单击“实例详情”可跳转至对应的在线服务详情页面。单击“预测”页签，进行服务测试。您也可以通过调用代码对服务进行测试，根据部署服务类型的不同，具体操作详情参见[访问在线服务](#)。

图 1-25 服务测试



下面的测试，是您在自动学习物体检测项目页面将模型部署上线之后进行服务测试的操作步骤。

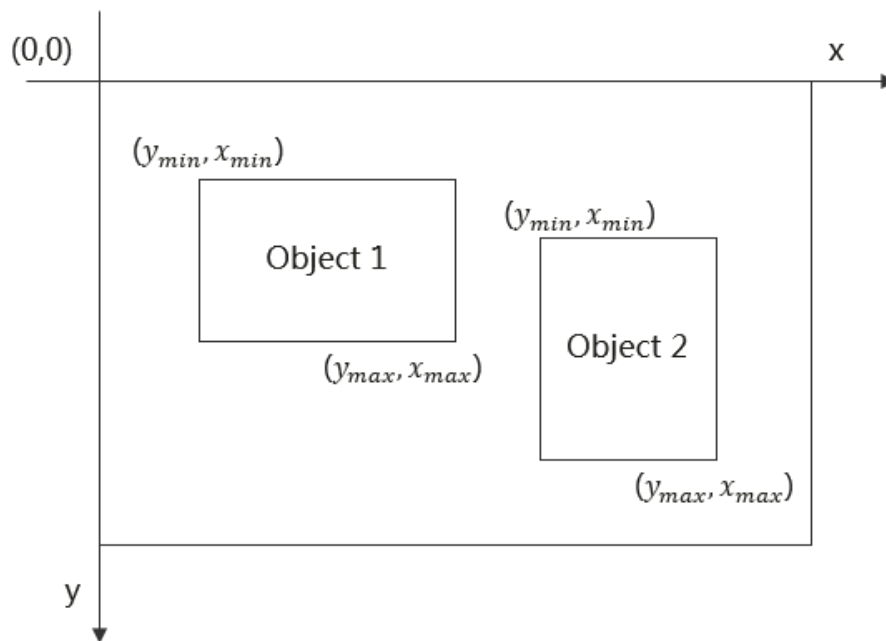
1. 模型部署完成后，“服务部署”节点，单击“实例详情”按钮，进入服务预测界面，在“预测”页签单击“上传”，选择本地图片进行测试。
2. 单击“预测”进行测试，预测完成后，右侧“预测结果”区域输出结果。如模型准确率不满足预期，可在“数据标注”页签中添加图片并进行标注，重新进行模型训练及部署上线。预测结果中的参数说明请参见表1-8。如果您对模型预测结果满意，可根据界面提示调用接口访问在线服务，操作指导请参见“[访问在线服务](#)”。

目前只支持jpg、jpeg、bmp、png格式的图片。

表 1-8 预测结果中的参数说明

参数	说明
detection_classes	每个检测框的标签。
detection_boxes	每个检测框的四点坐标 ($y_{min}, x_{min}, y_{max}, x_{max}$)，如图1-26所示。
detection_scores	每个检测框的置信度。

图 1-26 检测框的四点坐标示意图



📖 说明

- 由于“运行中”的在线服务将持续耗费资源，如果不需再使用此在线服务，建议在版本管理区域，单击“停止”，即可停止在线服务的部署，避免产生不必要的费用。如果需要继续使用此服务，可单击“启动”恢复。
- 如果您启用了自动停止功能，服务将在指定时间后自动停止，不再产生费用。

1.4 预测分析

1.4.1 准备数据

使用ModelArts自动学习构建预测分析模型时，您需要将数据上传至对象存储服务（OBS）中。OBS桶需要与ModelArts在同一区域，例如OBS桶区域为“北京四”时，必须保证ModelArts管理控制台区域也在“北京四”区域，否则会导致无法获取到相关数据。

数据集要求

预测分析项目中需要使用到的数据集为表格数据集，数据格式支持csv格式。表格数据集的具体介绍请参见[表格数据集简介](#)。

📖 说明

将原始.xlsx格式的数据转换为.csv格式的数据的方法如下：

将原始表格数据（.xlsx）另存。单击“文件>另存为”，选择本地地址后，下拉选择“保存类型”为“CSV（逗号分隔）(*.csv)”单击“保存”，在弹窗中，单击“确定”后就可以将.xlsx格式数据集转换为.csv格式。

表格数据集对训练数据的要求：

- 训练数据列数一致，总数据量不少于100条不同数据（有一个特征取值不同，即视为不同数据）。
- 训练数据列内容不能有时间戳格式（如：yy-mm-dd、yyyy-mm-dd等）的数据。
- 如果某一列的取值只有一种，会被视为无效列。请确保标签列的取值至少有两个且无数据缺失。

📖 说明

标签列指的是在训练任务中被指定为训练目标的列，即最终通过该数据集训练得到模型时的输出（预测项）。

- 除标签列外数据集中至少还应包含两个有效特征列（列的取值至少有两个且数据缺失比例低于10%）。
- 当前由于特征筛选算法限制，预测数据列建议放在数据集最后一列，否则可能导致训练失败。

表格数据集示例：

以银行存款预测数据集为例：根据预测人的年龄、工作类型、婚姻状况、文化程度、是否有房贷和是否有个人贷款。

表 1-9 数据源的具体字段及意义

字段名	含义	类型	描述
attr_1	年龄	Int	表示客户的年龄。
attr_2	职业	String	表示客户所从事的职业。
attr_3	婚姻情况	String	表示客户是否结婚或已离异。
attr_4	教育情况	String	表示客户受教育的程度。
attr_5	房产情况	String	表示客户名下是否有房产。
attr_6	贷款情况	String	表示客户名下是否有贷款。
attr_7	存款情况	String	表示客户名下是否有存款。

表 1-10 数据集样本数据

attr_1	attr_2	attr_3	attr_4	attr_5	attr_6	attr_7
31	blue-collar	married	secondary	yes	no	no
41	management	married	tertiary	yes	yes	no
38	technician	single	secondary	yes	no	no
39	technician	single	secondary	yes	no	yes
39	blue-collar	married	secondary	yes	no	no
39	services	single	unknown	yes	no	no

数据上传至 OBS

在本文档中，采用通过OBS管理控制台将数据上传至OBS桶。

上传OBS的文件规范：

预测分析项目的OBS数据路径需符合以下规则：

- 输入数据的OBS路径应指向数据文件，且文件不能直接放在OBS桶的根目录下，应该存放在OBS桶的文件夹内。如：“/obs-xxx/data/input.csv”。
- 输入数据的格式必须为csv格式，有效数据行数必须大于100行。列数必须小于200列，数据总大小不能超过100MB。

上传OBS操作步骤：

执行如下操作，将数据导入到您的数据集中，以便于模型训练和构建。

1. 登录OBS管理控制台，在ModelArts同一区域内[创建桶](#)。如果已存在可用的桶，需确保OBS桶与ModelArts在同一区域。
2. 参考[上传文件](#)，将本地数据上传至OBS桶中。如果您的数据较多，推荐OBS Browser+上传数据或上传文件夹。上传的数据需满足此类型自动学习项目的数据集要求。

📖 说明

在上传数据时，请选择非加密桶进行上传，否则会由于加密桶无法解密导致后期的训练失败。

创建数据集

数据准备完成后，需要创建预测分析项目支持的类型的表格数据集，具体操作请参考[创建数据集](#)。

常见问题

使用从OBS选择的数据创建表格数据集如何处理Schema信息？

Schema信息代表表格的列名和对应类型，需要跟导入数据的列数保持一致。

- 若您的原始表格中已包含表头，需要开启“导入是否包含表头”开关，系统会导入文件的第一行（表头）作为列名，无需再手动修改Schema信息。
- 若您的原始表格中没有表头，需关闭“导入是否包含表头”开关，从OBS选择数据后，Schema信息的列名默认为表格中的第一行数据，请更改Schema信息中的“列名”为attr_1、attr_2、……、attr_n，其中attr_n为最后一列，代表预测列。

1.4.2 创建项目

ModelArts自动学习，包括图像分类、物体检测、预测分析、声音分类和文本分类项目。您可以根据业务需求选择创建合适的项目。您需要执行如下操作来创建自动学习项目。

创建项目

1. 登录ModelArts管理控制台，在左侧导航栏单击“自动学习”，进入新版自动学习页面。
2. 在您需要的自动学习项目列表中。例如选择预测分析项目，单击“创建项目”，进入创建自动学习项目界面。
3. 在创建自动学习项目页面，计费模式默认“按需计费”，参考[表1-11](#)填写相应参数。

表 1-11 参数说明

参数	说明
“名称”	项目的名称。 <ul style="list-style-type: none">• 名称只能包含数字、字母、下划线和中划线，长度不能超过64位且不能为空。• 名称请以字母开头。• 名称不允许重复。


参数	说明
“描述”	对项目的简要描述。
“数据集”	<p>可在右侧下拉框选择已有数据集，或单击“创建数据集”前往新建数据集。</p> <ul style="list-style-type: none"> 已有数据集：在“数据集”右侧的下拉框中选择，仅展示同类型的数据集供选择。 创建数据集：前往创建数据集页面创建一个新的数据集。具体可参考如何创建数据集。
“标签列”	<p>可自行选择您需要预测的列名。</p> <p>标签列是预测模型的输出。模型训练步骤将使用全部信息训练预测模型，该模型以其他列的数据为输入，以标签列的预测值为输出。部署上线步骤将使用预测模型发布在线预测服务。</p>
“输出路径”	<p>选择自动学习数据输出的统一OBS路径。</p> <p>说明</p> <p>“输出路径”是存储自动学习在运行过程中所有产物的路径。</p>
“训练规格”	<p>选择自动学习训练节点所使用的资源规格，以实际界面显示为准，将会根据不同的规格计费。</p> <p>说明</p> <ul style="list-style-type: none"> 只有北京四区域支持限时免费规格。 若您购买了套餐包，可优先选择您对应规格的套餐包，在“配置费用”处会显示您的套餐余量，以及超出的部分如何计费，请您关注，避免造成不必要的资源浪费。

4. 单击“创建项目”，预测分析项目创建成功后页面自动跳转到“自动学习 workflow”。
5. 预测分析项目的工作流，将依次运行如下节点：
 - a. 数据集版本发布：将已完成确认的数据进行版本发布。
 - b. 数据校验：对您的数据集的数据进行校验，是否存在数据异常。
 - c. 预测分析：将发布好的数据集版本进行训练，生成对应的模型。
 - d. 模型注册：将训练后的结果注册到模型管理中。
 - e. 服务部署：将生成的模型部署为在线服务。

快速查找创建好的项目

在自动学习总览页，您可以通过搜索框，根据自动学习的属性类型（项目名称）快速搜索过滤到相应的工作流，可节省您的时间。


1. 登录ModelArts管理控制台，在左侧导航栏选择自动学习，进入自动学习总览页面。
2. 在自动学习列表上方的搜索框中，根据您需要的属性类型，例如，名称、状态、项目类型、当前节点、标签等，过滤出相应的工作流。

3. 单击搜索框右侧的  按钮，可选择自动学习的基础设置，需要的显示列。
表格内容折行：默认为关闭状态，启用此能力可让表格内容自动折行，禁用此功能可截断文本。

操作列：默认为关闭状态，启用此能力可让操作列固定在最后一列永久可见。
自定义显示列：默认所有显示项全部勾选，您可以根据实际需要定义您的显示列。

图 1-27 表格显示设置



- 单击“确定”即可按照设置好的显示列进行显示。
- 同时可支持对自动学习项目显示页进行排序，单击表头中的箭头，就可对该列进行排序。

1.4.3 模型训练

创建自动学习后，将会进行模型的训练，得到预测分析的模型。部署上线步骤将使用预测模型发布在线预测服务。

操作步骤


- 在新版自动学习页面，单击创建成功的项目名称，查看当前工作流的执行情况。
- 在“预测分析”节点中，待节点状态由“运行中”变为“运行成功”，即完成了模型的自动训练。
- 训练完成后，您可以在预测分析节点中单击查看训练详情，如“标签列”和“标签列数据类型”、“准确率”、“评估结果”等。
该示例为二分类的离散型数值，评估效果参数说明请参见[表1-12](#)。
不同类型标签列数据产生的评估结果说明请参见[评估结果说明](#)。

图 1-28 模型评估报告



Metric	Value
accuracy	0.93
balanced_accuracy	0.77
f1	0.5
mcc	0.509
precision	1
recall	0.33
roc_auc	0.952

说明

同一个自动学习项目可以训练多次，每次训练会注册一个新的AI应用一个版本。如第一次训练版本号“0.0.1”，下一个版本为“0.0.2”。基于训练版本可以对训练模型进行管理。当训练的模型达到目标后，再执行部署上线的操作。

评估结果说明

根据训练数据类的不同评估结果会包含不同的指标。

- 离散值评估结果
包含评估指标为召回率（Recall）、精确率（Precision）、准确率（Accuracy）与F1值（F1 Score）。下表为具体说明：

表 1-12 离散值评估结果包含指标说明

参数	说明
recall：召回率	被用户标注为某个分类的所有样本中，模型正确预测为该分类的样本比率，反映模型对正样本的识别能力。
precision：精确率	被模型预测为某个分类的所有样本中，模型正确预测的样本比率，反映模型对负样本的区分能力。
accuracy：准确率	所有样本中，模型正确预测的样本比率，反映模型对样本整体的识别能力。
f1：F1值	F1值是模型精确率和召回率的加权调和平均，用于评价模型的好坏，当F1较高时说明模型效果较好。

- 连续数值评估结果
包含评估指标为平均绝对误差（Mean Absolute Error）、均方误差（Mean Squared Error）与均方根误差（Root Mean Squared Error）。三个误差值能够表征真实值和预测值之间的差距。在多次建模的过程中，每一次建模结果都会产生一组误差值，评判一个模型好坏的方法就是看这三个误差值是否变小或者变大，误差值越小表示模型越好。

1.4.4 部署上线

部署上线

部署上线操作即将模型部署为在线服务，并且提供在线的测试UI与监控能力。完成模型训练后，可选择准确率理想且训练状态为“运行成功”的版本部署上线。具体操作步骤如下。

1. 在“运行节点”页面中，待训练状态变为“等待输入”，双击“服务部署”节点，完成相关参数配置。
2. 在服务部署页面，选择部署上线使用的资源规格。
 - AI应用来源：默认为生成的AI应用。
 - 选择AI应用及版本：自动匹配当前使用的AI应用版本，支持选择版本。
 - 资源池：默认公共资源池。
 - 分流：默认为100，输入值必须是0-100之间。
 - 计算节点规格：请根据界面显示的列表，选择可用的规格，置灰的规格表示当前环境无法使用。如果公共资源池下规格为空数据，表示当前环境无公共资源。建议使用专属资源池，或者联系系统管理员创建公共资源池。
 - 计算节点个数：默认为1，输入值必须是1-5之间的整数。
 - 是否自动停止：启用该参数并设置时间后，服务将在指定时间后自动停止。如果不启用此参数，在线服务将一直运行，同时一直收费，自动停止功能可以帮您避免产生不必要的费用。默认开启自动停止功能，且默认值为“1小时后”。

目前支持设置为“1小时后”、“2小时后”、“4小时后”、“6小时后”、“自定义”。如果选择“自定义”的模式，可在右侧输入框中输入1~24范围内的任意整数。

说明

若您购买了套餐包，计算节点规格可选择您的套餐包，同时在“配置费用”页签还可查看您的套餐包余量以及超出部分的计费方式，请您务必关注，避免造成不必要的资源浪费。

3. 完成资源配置后，单击“继续运行”，在弹框中确认继续运行后，服务部署节点将继续运行，直至状态变为“运行成功”，至此，已将AI应用部署为在线服务。

服务测试

服务部署节点运行成功后，单击“实例详情”可跳转至对应的在线服务详情页面。单击“预测”页签，进行服务测试。您也可以通过调用代码对服务进行测试，根据部署服务类型的不同，具体操作详情参见[访问在线服务](#)、[访问边缘服务](#)。

图 1-29 服务测试



下面的测试，是您在自动学习预测分析项目页面将模型部署上线之后进行服务测试的操作步骤。

1. 模型部署完成后，您可输入代码进行测试。在“自动学习”页面，在服务部署节点，单击“实例详情”进入“在线服务”界面，在“预测”页签的“预测代码”区域，输入调试代码。
2. 单击“预测”进行测试，预测完成后，右侧“返回结果”区域输出测试结果。如模型准确率不满足预期，可在“数据标注”页签，重新进行模型训练及部署上线。如果您对模型预测结果满意，可根据界面提示调用接口访问在线服务，操作指导请参见“[访问在线服务](#)”。
 - 输入代码：其中预测分析要求数据集中数据的预测列名称为class，否则会导致预测失败。

```
{
  "data": {
    "req_data": [{
      "attr_1": "34",
      "attr_2": "blue-collar",
      "attr_3": "single",
      "attr_4": "tertiary",
      "attr_5": "no",
      "attr_6": "tertiary"
    }]
  }
}
```

- 返回结果如[图1-30](#)所示：predict为目标列的预测结果。

图 1-30 预测结果

```
1 {
2   "data": {
3     "req_data": [{
4       "attr_1": "34",
5       "attr_2": "blue-collar",
6       "attr_3": "single",
7       "attr_4": "tertiary",
8       "attr_5": "no",
9       "attr_6": "tertiary"
10    }]
11  }
12 }
```

```
1 {
2   "data": {
3     "resp_data": [
4       {
5         "predict": "no"
6       }
7     ]
8   }
9 }
```

📖 说明

- 由于“运行中”的在线服务将持续耗费资源，如果不需再使用此在线服务，建议在“在线服务”的操作列表单击“更多>停止”，即可停止在线服务的部署，避免产生不必要的费用。如果需要继续使用此服务，可单击“启动”恢复。
- 如果您启用了自动停止功能，服务将在指定时间后自动停止，不再产生费用。

1.5 声音分类

1.5.1 准备数据

使用ModelArts自动学习构建模型时，您需要将数据上传至对象存储服务（OBS）中。OBS桶需要与ModelArts在同一区域。

声音分类的数据要求

- 音频只支持16bit的WAV格式。支持WAV的所有子格式。
- 单条音频时长应大于1s，大小不能超过4MB。
- 适当增加训练数据，会提升模型的精度。声音分类建议每类音频至少20条，每类音频总时长至少5分钟。
- 建议训练数据和真实识别场景的声音保持一致并且每类的音频尽量覆盖真实环境的所有场景。

- 训练集的数据质量对于模型的精度有很大影响，建议训练集音频的采样率和采样精度保持一致。
- 标注质量对于最终的模型精度有极大的影响，标注过程中尽量不要出现误标情况。
- 音频标注涉及到的标注标签和声音内容只支持中文和英文，不支持小语种。

数据上传至 OBS

在本文档中，采用通过OBS管理控制台将数据上传至OBS桶。

上传OBS的文件规范：

- 如不需要提前上传训练数据，请创建一个空文件夹用于存放工程后期生成的文件。如：“/bucketName/data-cat”。
- 如需要提前上传待标注的音频，请创建一个空文件夹，然后将音频文件保存在该文件夹下，音频的目录结构如：“/bucketName/data-cat/cat.wav”。

上传OBS的操作步骤：

执行如下操作，将数据导入到您的数据集中，以便用于模型训练和构建。

1. 登录OBS管理控制台，在ModelArts同一区域内[创建桶](#)。如果已存在可用的桶，需确保OBS桶与ModelArts在同一区域。
2. 参考[上传文件](#)，将本地数据上传至OBS桶中。如果您的数据较多，推荐OBS Browser+上传数据或上传文件夹。上传的数据需满足此类型自动学习项目的数据集要求。

说明

- 在上传数据时，请选择非加密桶进行上传，否则会由于加密桶无法解密导致后期的训练失败。
- 用于训练的音频，至少有2种以上的分类，每种分类的音频数据数不少20条。

创建数据集

数据准备完成后，需要创建相应项目支持的类型的数据集，具体操作请参考[创建数据集](#)。

1.5.2 创建项目

ModelArts自动学习，包括图像分类、物体检测、预测分析、声音分类和文本分类项目。您可以根据业务需求选择创建合适的项目。您需要执行如下操作来创建自动学习项目。

创建项目

1. 登录ModelArts管理控制台，在左侧导航栏单击“自动学习”，进入新版自动学习页面。
2. 在您需要的自动学习项目列表中，单击“创建项目”，进入创建自动学习项目界面。
3. 在创建自动学习项目页面，计费模式默认“按需计费”，参考[表1-13](#)填写相应参数。

表 1-13 参数说明

参数	说明
“名称”	项目的名称。 <ul style="list-style-type: none"> 名称只能包含数字、字母、下划线和中划线，长度不能超过64位且不能为空。 名称请以字母开头。 名称不允许重复。
“描述”	对项目的简要描述。
“数据集”	可在右侧下拉框选择已有数据集，或单击“创建数据集”前往新建数据集。 <ul style="list-style-type: none"> 已有数据集：在“数据集”右侧的下拉框中选择，仅展示同类型的数据集供选择。 创建数据集：前往创建数据集页面创建一个新的数据集。具体可参考如何创建数据集。
“输出路径”	选择自动学习数据输出的统一OBS路径。 说明 “输出路径”是存储自动学习在运行过程中所有产物的路径。
“训练规格”	选择自动学习训练节点所使用的资源规格，以实际界面显示为准，将会根据不同的规格计费。 说明 <ul style="list-style-type: none"> 只有北京四区域支持限时免费规格。 若您购买了套餐包，可优先选择您对应规格的套餐包，在“配置费用”处会显示您的套餐余量，以及超出的部分如何计费，请您关注，避免造成不必要的资源浪费。

4. 单击“创建项目”，声音分类项目创建成功后页面自动跳转到“自动学习 workflow”。
5. 声音分类项目的工作流，将依次运行如下节点：
 - a. 数据标注：对您的数据进行标注情况确认。
 - b. 数据集版本发布：将已完成确认的数据进行版本发布。
 - c. 数据校验：对您的数据集的数据进行校验，是否存在数据异常。
 - d. 声音分类：将发布好的数据集版本进行训练，生成对应的模型。
 - e. 模型注册：将训练后的结果注册到模型管理中。
 - f. 服务部署：将生成的模型部署为在线服务。

快速查找创建好的项目

在自动学习总览页，您可以通过搜索框，根据自动学习的属性类型（项目名称）快速搜索过滤到相应的工作流，可节省您的时间。

1. 登录ModelArts管理控制台，在左侧导航栏选择自动学习，进入自动学习总览页面。
2. 在自动学习列表上方的搜索框中，根据您需要的属性类型，例如，名称、状态、项目类型、当前节点、标签等，过滤出相应的工作流。



- 单击搜索框右侧的  按钮，可选择自动学习的基础设置，需要的显示列。
表格内容折行：默认为关闭状态，启用此能力可让表格内容自动折行，禁用此功能可截断文本。
操作列：默认为关闭状态，启用此能力可让操作列固定在最后一列永久可见。
自定义显示列：默认所有显示项全部勾选，您可以根据实际需要定义您的显示列。

图 1-31 表格显示设置



- 单击“确定”即可按照设置好的显示列进行显示。
- 同时可支持对自动学习项目显示页进行排序，单击表头中的箭头 ，就可对该列进行排序。


1.5.3 数据标注

项目创建完成后，将会自动跳转至新版自动学习页面，并开始运行，当数据标注节点的状态变为“等待操作”时，需要手动进行确认数据集中的数据标注情况，也可以对数据集中的数据进行标签的修改，数据的增加或删减。

图 1-32 数据标注节点状态

数据标注

运行状况

 请前往 [实例详情](#) 页面进行标注

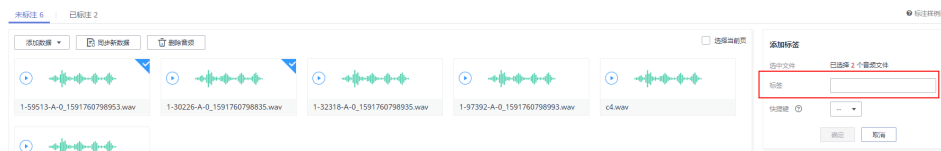
属性

状态 ● 等待操作

音频标注

1. 在新版自动学习页面单击“实例详情”按钮，前往数据标注页面。单击任意一张图片，进入音频标注页面。
2. 在“音频标注”页面单击“未标注”页签，此页面展示所有未标注的音频数据。依次单击选中待标注的音频，或勾选“选择当前页”选中该页面所有音频，在页面右侧进行标注。

图 1-33 音频标注



3. 添加标注。先对音频进行播放识别，然后选中音频文件，在右侧“标签”区域，输入“标签名”或从下拉列表中选择已添加的标签，同时可在下拉菜单中选择标签“快捷键”。单击“确定”，完成选中音频的标注操作。
4. 当目录中所有音频都完成标注后，您可以在“已标注”页签下查看已完成标注的音频，或者通过右侧的“全部标签”列表，了解当前已完成的标签名称和标签数量。

同步或添加音频

在“数据标注”节点单击“实例详情”进入“音频标注”页面。声音分类项目创建时，音频来源有两种，通过本地添加或同步OBS中的数据。

- **添加音频：**您可以将本地音频快速添加到ModelArts，同时自动上传至创建项目时所选择的OBS路径中。单击“添加数据”，在弹出的对话框中输入正确的数据并添加。

说明

仅支持16bit WAV格式音频文件，单个音频文件不能超过4MB，且单次上传的音频文件总大小不能超过8MB。

- **数据源同步：**为了快速获取用户OBS桶中最新音频，单击“数据源同步”，快速将通过OBS上传的音频数据添加到ModelArts。

- **删除音频：**您可以依次单击选中音频，或勾选“选择当前页”选中该页面所有音频进行删除操作。

📖 说明

所有的删除操作均不可恢复，请谨慎操作。

修改标注

当数据完成标注后，您还可以进入“已标注”页签，对已标注的数据进行修改。

- **基于音频修改**

在数据集详情页，单击“已标注”页签，然后在音频列表中选中待修改的音频（选择一个或多个）。在右侧标签信息区域中对标签进行修改。

- **修改标签：**在“选中文件标签”区域中，单击操作列的编辑图标，然后在文本框中输入正确的标签名，然后单击确定图标完成修改。
- **删除标签：**在“选中文件标签”区域中，单击操作列的删除图标，在弹出的对话框中单击“确定”删除该标签。

- **基于标签修改**

在数据标注页面，单击右侧的“标签管理”，在标签管理页，显示全部标签的信息。

- **修改标签：**单击操作列的“修改”按钮，在弹出的对话框中输入修改后的标签名、选择修改后的快捷键，然后单击“确定”完成修改。修改后，之前添加了此标签的音频，都将被标注为新的标签名称。
- **删除标签：**单击操作列的“删除”按钮，在弹出的对话框中，根据提示选择删除对象，然后单击“确定”。

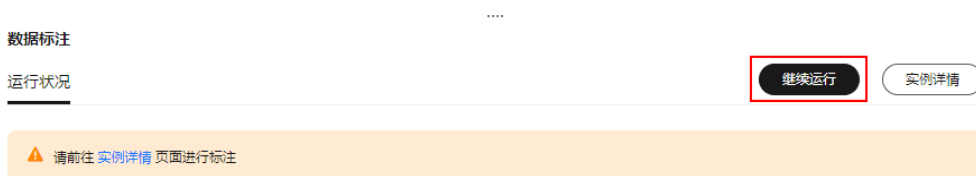
📖 说明

删除后的标签无法恢复，请谨慎操作。

继续运行

完成数据的确认之后，返回新版自动学习的页面，在数据标注节点单击“继续运行”， workflow 将会继续依次运行直到所有节点运行成功。

图 1-34 继续运行



1.5.4 模型训练

完成音频标注后，可以进行模型的训练。模型训练的目的在于得到满足需求的声音分类模型。由于用于训练的音频，至少有2种以上的分类，每种分类的音频数不少于5个。

操作步骤

在开始训练之前，需要完成数据标注，然后再开始模型的自动训练。


1. 在新版自动学习页面，单击项目名称进入运行总览页面，单击数据标注节点的“实例详情”进入数据标注页面，完成数据标注。
2. 返回新版自动学习页面，单击数据标注节点的“继续运行”，然后等待 workflow 按顺序进入训练节点。
3. 模型将会自动进入训练，无需人工介入，训练时间相对较长，建议您耐心等待。如果关闭或退出此页面，系统仍然在执行训练操作。
4. 在“声音分类”节点中，待训练状态由“运行中”变为“运行成功”，即完成模型的自动训练。
5. 训练完成后，您可以单击声音分类节点上方的按钮，查看相关指标信息，如“准确率”、“评估结果”等。

表 1-14 评估结果参数说明

参数	说明
recall: 召回率	被用户标注为某个分类的所有样本中，模型正确预测为该分类的样本比率，反映模型对正样本的识别能力。
precision: 精确率	被模型预测为某个分类的所有样本中，模型正确预测的样本比率，反映模型对负样本的区分能力。
accuracy: 准确率	所有样本中，模型正确预测的样本比率，反映模型对样本整体的识别能力。
f1: F1值	F1值是模型精确率和召回率的加权调和平均，用于评价模型的好坏，当F1较高时说明模型效果较好。

📖 说明

同一个自动学习项目可以训练多次，每次训练会注册一个新的AI应用版本。如第一次训练版本号为“0.0.1”，下一个版本为“0.0.2”。基于训练版本可以对训练模型进行管理。当训练的模型达到目标后，再执行部署上线的操作。

1.5.5 部署上线

部署上线

部署上线操作即将模型部署为在线服务，并且提供在线的测试UI与监控能力。完成模型训练后，可选择准确率理想且训练状态为“运行成功”的版本部署上线。具体操作步骤如下。

1. 在“运行总览”页面中，待服务部署节点的状态变为“等待输入”时，双击“服务部署”进入配置详情页，完成资源的参数配置操作。
2. 在服务部署页面，选择部署上线使用的资源规格。
 - AI应用来源：默认为生成的AI应用。
 - 选择AI应用及版本：自动匹配当前使用的AI应用版本，支持选择版本。
 - 资源池：默认公共资源池。
 - 分流：默认为100，输入值必须是0-100之间。

- 计算节点规格：请根据界面显示的列表，选择可用的规格，置灰的规格表示当前环境无法使用。如果公共资源池下规格为空数据，表示当前环境无公共资源。建议使用专属资源池，或者联系系统管理员创建公共资源池。
- 计算节点个数：默认为1，输入值必须是1-5之间的整数。
- 是否自动停止：启用该参数并设置时间后，服务将在指定时间后自动停止。如果不启用此参数，在线服务将一直运行，同时一直收费，自动停止功能可以帮您避免产生不必要的费用。默认开启自动停止功能，且默认值为“1小时后”。
目前支持设置为“1小时后”、“2小时后”、“4小时后”、“6小时后”、“自定义”。如果选择“自定义”的模式，可在右侧输入框中输入1~24范围内的任意整数。

说明

若您购买了套餐包，计算节点规格可选择您的套餐包，同时在“配置费用”页签还可查看您的套餐包余量以及超出部分的计费方式，请您务必关注，避免造成不必要的资源浪费。

3. 完成资源配置后，单击“继续运行”，在弹框中确认继续运行后，服务部署节点将继续运行，直至状态变为“运行成功”，至此，已将AI应用部署为在线服务。

服务测试

服务部署节点运行成功后，单击“实例详情”可跳转至对应的在线服务详情页面。单击“预测”页签，进行服务测试。您也可以通过调用代码对服务进行测试，根据部署服务类型的不同，具体操作详情参见[访问在线服务](#)、[访问边缘服务](#)。

图 1-35 服务测试



下面的测试，是您在自动学习声音分类项目页面将模型部署上线之后进行服务测试的操作步骤。

1. 模型部署完成后，您可添加音频文件进行测试。在“自动学习”页面，选择服务部署节点，单击实例详情，进入“部署上线”界面，选择状态为“运行中”的服务版本，在“服务测试”区域单击“上传”，选择本地音频进行测试。
2. 单击“预测”进行测试，预测完成后，右侧“预测结果”区域输出测试结果。如模型准确率不满足预期，可在“数据标注”页签中添加音频并进行标注，重新进行模型训练及部署上线。预测结果中的参数说明请参见表1-15。如果您对模型预测结果满意，可根据界面提示调用接口访问在线服务，操作指导请参见[“访问在线服务”](#)。

表 1-15 预测结果中的参数说明

参数	说明
predicted_label	该段音频的预测类别。
score	预测为此类别的置信度。

📖 说明

由于“运行中”的在线服务将持续耗费资源，如果不再使用此在线服务，建议在版本管理区域，单击“停止”，即可停止在线服务的部署，避免产生不必要的费用。如果需要继续使用此服务，可单击“启动”恢复。

如果您启用了自动停止功能，服务将在指定时间后自动停止，不再产生费用。

1.6 文本分类

1.6.1 准备数据

使用ModelArts自动学习构建模型时，您需要将数据上传至对象存储服务（OBS）中。OBS桶需要与ModelArts在同一区域。

数据集要求

- 文件格式要求为txt或者csv，文件大小不能超过8MB。
- 以换行符作为分隔符，每行数据代表一个标注对象。
- 文本分类目前只支持中文。

数据上传至 OBS

在本文档中，采用通过OBS管理控制台将数据上传至OBS桶。

OBS上传文件的规范：

- 如不需要提前上传训练数据，请创建一个空文件夹用于存放工程后期生成的文件。
- 如需要提前上传待标注的文件，请创建一个空文件夹，然后将文本文件保存在该文件夹下，文本文件的目录结构如：“/bucketName/data/text.csv”。
- 标签名是由中文、大小写字母、数字、中划线或下划线组成，且不超过32位的字符串。
- 如您将已标注好的文本文件上传至OBS桶，请按照如下规范上传。
 - 要求将标注对象和标注文件存储在同一目录，并且一一对应，如标注对象文件名为“COMMENTS_114745.txt”，那么标注文件名为“COMMENTS_114745_result.txt”。

数据文件存储示例：

```
|<dataset-import-path>
|
|   COMMENTS_114732.txt
|   COMMENTS_114732_result.txt
|   COMMENTS_114745.txt
|   COMMENTS_114745_result.txt
|   COMMENTS_114945.txt
|   COMMENTS_114945_result.txt
```

- 文本分类的标注对象和标注文件均为文本文件，并且以行数进行对应。如标注文件中的第一行表示的是标注对象文件中的第一行的标注。

例如，标注对象“COMMENTS_114745.txt”的内容如下所示。

```
手感很好，反应速度很快，不知道以后怎样
三个月前买了一个用的非常好果断把旧手机替换下来尤其在待机方面
没充一会电源怎么也会发热呢音量键不好用回弹不好
算是给自己的父亲节礼物吧物流很快下单不到24小时就到货了耳机更赞有些低音炮的感觉入耳很紧不会掉棒棒哒
```

标注文件“COMMENTS_114745_result.txt”的内容。

```
positive  
positive  
negative  
positive
```

OBS上传操作步骤：

执行如下操作，将数据导入到您的数据集中，以便用于模型训练和构建。

1. 登录OBS管理控制台，在ModelArts同一区域内[创建桶](#)。如果已存在可用的桶，需确保OBS桶与ModelArts在同一区域。
2. 参考[上传文件](#)，将本地数据上传至OBS桶中。如果您的数据较多，推荐OBS Browser+上传数据或上传文件夹。上传的数据需满足此类型自动学习项目的数据集要求。

📖 说明

- 在上传数据时，请选择非加密桶进行上传，否则会由于加密桶无法解密导致后期的训练失败。
- 用于训练的文本，至少有2种以上的分类，每种分类样本数据数不少20行。

创建数据集

数据准备完成后，需要创建相应项目支持的类型的数据集，具体操作请参考[创建数据集](#)。

1.6.2 创建项目

ModelArts自动学习，包括图像分类、物体检测、预测分析、声音分类和文本分类项目。您可以根据业务需求选择创建合适的项目。您需要执行如下操作来创建自动学习项目。

创建项目

1. 登录ModelArts管理控制台，在左侧导航栏单击“自动学习”，进入新版自动学习页面。
2. 在您需要的自动学习项目列表中，单击“创建项目”，进入创建自动学习项目界面。
3. 在创建自动学习项目页面，计费模式默认“按需计费”，参考[表1-16](#)填写相应参数。

表 1-16 参数说明


参数	说明
“名称”	项目的名称。 <ul style="list-style-type: none">• 名称只能包含数字、字母、下划线和中划线，长度不能超过64位且不能为空。• 名称请以字母开头。• 名称不允许重复。
“描述”	对项目的简要描述。

参数	说明
“数据集”	<p>可在右侧下拉框选择已有数据集，或单击“创建数据集”前往新建数据集。</p> <ul style="list-style-type: none"> • 已有数据集：在“数据集”右侧的下拉框中选择，仅展示同类型的数据集供选择。 • 创建数据集：前往创建数据集页面创建一个新的数据集。具体可参考如何创建数据集。
“输出路径”	<p>选择自动学习数据输出的统一OBS路径。</p> <p>说明 “输出路径”是存储自动学习在运行过程中所有产物的路径。</p>
“训练规格”	<p>选择自动学习训练节点所使用的资源规格，以实际界面显示为准，将会根据不同的规格计费。</p> <p>说明</p> <ul style="list-style-type: none"> • 只有北京四区域支持限时免费规格。 • 若您购买了套餐包，可优先选择您对应规格的套餐包，在“配置费用”处会显示您的套餐余量，以及超出的部分如何计费，请您关注，避免造成不必要的资源浪费。

4. 单击“创建项目”，预测分析项目创建成功后页面自动跳转到“自动学习 workflow”。
5. 文本分类项目的工作流，将依次运行如下节点：
 - a. 数据标注：对您的数据进行标注情况确认。
 - b. 数据集版本发布：将已完成确认的数据进行版本发布。
 - c. 数据校验：对您的数据集的数据进行校验，是否存在数据异常。
 - d. 文本分类：将发布好的数据集版本进行训练，生成对应的模型。
 - e. 模型注册：将训练后的结果注册到模型管理中。
 - f. 服务部署：将生成的模型部署为在线服务。

快速查找创建好的项目

在自动学习总览页，您可以通过搜索框，根据自动学习的属性类型（项目名称）快速搜索过滤到相应的工作流，可节省您的时间。

1. 登录ModelArts管理控制台，在左侧导航栏选择自动学习，进入自动学习总览页面。
2. 在自动学习列表上方的搜索框中，根据您需要的属性类型，例如，名称、状态、项目类型、当前节点、标签等，过滤出相应的工作流。
3. 单击搜索框右侧的  按钮，可选择自动学习的基础设置，需要的显示列。


表格内容折行：默认为关闭状态，启用此能力可让表格内容自动折行，禁用此功能可截断文本。

操作列：默认为关闭状态，启用此能力可让操作列固定在最后一列永久可见。

自定义显示列：默认所有显示项全部勾选，您可以根据实际需要定义您的显示列。

图 1-36 表格显示设置



4. 单击“确定”即可按照设置好的显示列进行显示。
5. 同时可支持对自动学习项目显示页进行排序，单击表头中的箭头，就可对该列进行排序。

1.6.3 数据标注

项目创建完成后，将会自动跳转至新版自动学习页面，并开始运行，当数据标注节点的状态变为“等待操作”时，需要手动进行确认数据集中的数据标注情况，也可以对数据集中的数据进行标签的修改，数据的增加或删减。

图 1-37 数据标注节点状态

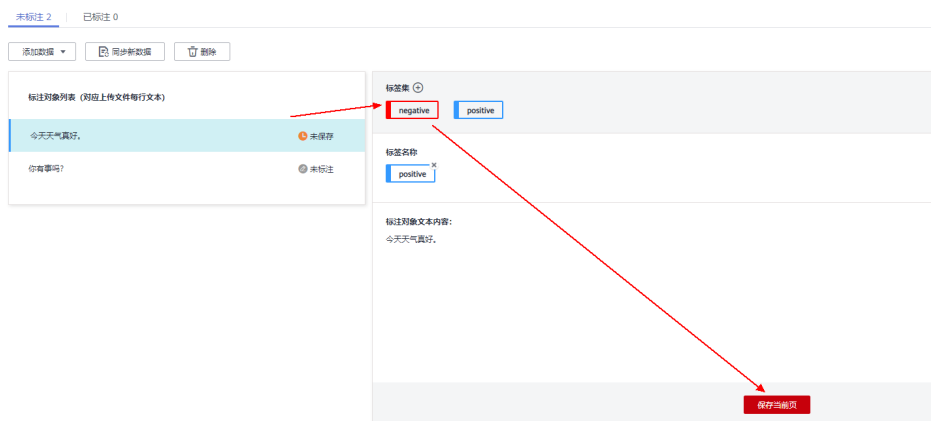


双击“数据标注”节点，单击**实例详情**按钮，打开数据标注页面。

文本分类的数据标注

1. 在“标注对象列表”中选中待标注文本，然后单击“标签集”区域中不同标签进行标注。
一个文本对象只能添加一个标签。
2. 确认文件标签后，单击右下方“保存当前页”，完成标注。
当“标注对象列表”内容较多时，其区域下方将呈现翻页，请务必在本页完成标注后，单击“保存当前页”保存后再翻页。如果您未完成保存即翻页，将导致前一页的标注信息丢失，需重新标注。

图 1-38 数据标注-文本分类



添加或删除数据

自动学习项目中，数据来源为数据集中输入位置对应的OBS目录，当目录下的数据无法满足现有业务时，您可以在ModelArts自动学习页面中，添加或删除数据。

- **添加文件**
在“未标注”页签下，可单击页面左上角的“添加数据”，您可以在弹出对话框中，选择本地文件上传。
上传文件格式需满足文本分类型的[数据集要求](#)。
- **删除文本对象**
在“已标注”页签或“未标注”页签下，选中需要删除的文本对象，单击页面左上角的“删除”，在弹出的对话框中，确认删除信息后，单击“确定”。
在“已标注”页签下，您还可以勾选“选择当前页”，单击“删除”，即可删除当前页下所有的文本对象及其标注信息。

修改已标注的数据

针对“已标注”的文本数据，仅支持删除此文本对象的标签。在“已标注”页签下，在标签名称区域单击标签右上角的叉号，即可删除此文本对象的标签。标签删除后，此文本对象将被呈现至“未标注”页签下。

图 1-39 删除已标注文本的标签



修改标签

针对文本分类的自动学习项目，项目创建成功后，您可以根据业务变化，修改用于标注的标签。支持添加、修改和删除标签。

- 添加标签
在“未标注”页签下，单击“标签集”右侧的加号，在弹出“新增标签”对话框中，设置“标签名称”和“标签颜色”，然后单击“确定”完成标签添加。
- 修改标签
在“已标注”页签中“全部标签”的下方操作列，选择需要修改的标签，单击操作列的编辑图标，在弹出“修改标签”对话框中，修改“标签名称”或“标签颜色”，然后单击“确定”完成标签修改。
- 删除标签
在“已标注”页签中“全部标签”的下方，选择需要删除的标签，单击操作列的删除图标，在弹出“删除”对话框中，选择“仅删除标签”或“删除标签及仅包含此标签的标注对象”，然后单击“确定”完成标签删除。

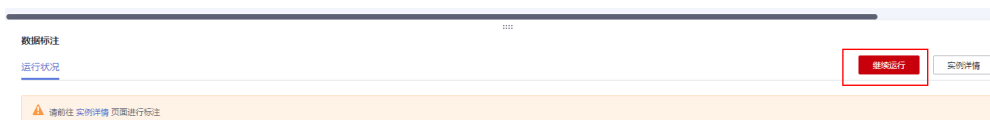
说明

所有的删除操作均不可恢复，请谨慎操作。

继续运行

完成数据的确认之后，返回新版自动学习的页面，在数据标注节点单击“继续运行”， workflows 将会继续依次运行直到所有节点运行成功。

图 1-40 继续运行



1.6.4 模型训练

完成数据标注后，可进行模型的训练。模型训练的目的在于得到满足需求的文本分类模型。由于用于训练的文本，至少有2种以上的分类（即2种以上的标签），每种分类的

文本数不少于20个。因此在单击“继续运行”按钮之前，请确保已标注的文本符合要求。

操作步骤

1. 在新版自动学习页面，单击项目名称进入运行总览，单击“数据标注”节点的“实例详情”进入“数据标注”页面，完成数据标注。

图 1-41 完成数据标注



2. 返回新版自动学习页面，单击数据标注节点的“继续运行”，然后等待 workflow 按顺序进入训练节点。
3. 模型将会自动进入训练，无需人工介入，训练时间相对较长，建议您耐心等待。如果关闭或退出此页面，系统仍然在执行训练操作。
4. 在“文本分类”节点中，待训练状态由“运行中”变为“运行成功”，即完成模型的自动训练。

图 1-42 运行成功




5. 训练完成后，您可以单击文本分类节点上方的  按钮，查看相关指标信息，如“准确率”、“评估结果”等。评估结果参数说明请参见[表1-17](#)。

图 1-43 模型评估报告



Metric	Value
accuracy	0.93
balanced_accuracy	0.77
f1	0.5
mcc	0.509
precision	1
recall	0.33
roc_auc	0.952

表 1-17 评估结果参数说明

参数	说明
recall: 召回率	被用户标注为某个分类的所有样本中，模型正确预测为该分类的样本比率，反映模型对正样本的识别能力。
precision: 精确率	被模型预测为某个分类的所有样本中，模型正确预测的样本比率，反映模型对负样本的区分能力。
accuracy: 准确率	所有样本中，模型正确预测的样本比率，反映模型对样本整体的识别能力。
f1: F1值	F1值是模型精确率和召回率的加权调和平均，用于评价模型的好坏，当F1较高时说明模型效果较好。

📖 说明

同一个自动学习项目可以训练多次，每次训练生成一个版本。如第一次训练版本号为“0.0.1”，下一个版本为“0.0.2”。基于训练版本可以对训练模型进行管理。当训练的模型达到目标后，再执行部署上线的操作。

1.6.5 部署上线

部署上线

部署上线操作即将模型部署为在线服务，并且提供在线的测试UI与监控能力。完成模型训练后，可选择准确率理想且训练状态为“运行成功”的版本部署上线。具体操作步骤如下。

1. 在“运行总览”页面中，待服务部署节点的状态变为“等待输入”，双击“服务部署”节点，进入配置详情页，完成资源的参数配置操作。
2. 在服务部署页面，选择部署上线使用的资源规格。
 - AI应用来源：默认为生成的AI应用。
 - 选择AI应用版本：自动匹配当前使用的AI应用版本，支持选择版本。
 - 资源池：默认公共资源池。

- 分流：默认为100，输入值必须是0-100之间。
- 计算节点规格：请根据界面显示的列表，选择可用的规格，置灰的规格表示当前环境无法使用。如果公共资源池下规格为空数据，表示当前环境无公共资源。建议使用专属资源池，或者联系系统管理员创建公共资源池。
- 计算节点个数：默认为1，输入值必须是1-5之间的整数。
- 是否自动停止：启用该参数并设置时间后，服务将在指定时间后自动停止。如果不启用此参数，在线服务将一直运行，同时一直收费，自动停止功能可以帮您避免产生不必要的费用。默认开启自动停止功能，且默认值为“1小时后”。
目前支持设置为“1小时后”、“2小时后”、“4小时后”、“6小时后”、“自定义”。如果选择“自定义”的模式，可在右侧输入框中输入1~24范围内的任意整数。

📖 说明

若您购买了套餐包，计算节点规格可选择您的套餐包，同时在“配置费用”页签还可查看您的套餐包余量以及超出部分的计费方式，请您务必关注，避免造成不必要的资源浪费。

3. 完成资源配置后，单击“继续运行”，在弹框中确认继续运行后，服务部署节点将继续运行，直至状态变为“运行成功”，至此，已将AI应用部署为在线服务。

服务测试

服务部署节点运行成功后，单击“实例详情”可跳转至对应的在线服务详情页面。单击“预测”页签，进行服务测试。您也可以通过调用代码对服务进行测试，根据部署服务类型的不同，具体操作详情参见[访问在线服务](#)、[访问边缘服务](#)。

图 1-44 服务测试



下面的测试，是您在自动学习文本分类项目页面将模型部署上线之后进行服务测试的操作步骤。

1. 模型部署完成后，您可添加文本进行测试。在“自动学习”页面，选择目标项目，进入“部署上线”界面，选择状态为“运行中”的服务版本，在“服务测试”区域的文本框中，输入需测试的文本。
2. 单击“预测”进行测试，预测完成后，右侧“预测结果”区域输出测试结果。如模型准确率不满足预期，可在“数据标注”页签中添加数据并进行标注，重新进行模型训练及部署上线。预测结果中的参数说明请参见表1-18。如果您对模型预测结果满意，可根据界面提示调用接口访问在线服务，操作指导请参见[“访问在线服务”](#)。

表 1-18 预测结果中的参数说明

参数	说明
predicted_label	该段文本的预测类别。

参数	说明
score	预测为此类别的置信度。

📖 说明

由于“运行中”的在线服务将持续耗费资源，如果不再使用此在线服务，建议在版本管理区域，单击“停止”，即可停止在线服务的部署，避免产生不必要的费用。如果需要继续使用此服务，可单击“启动”恢复。

如果您启用了自动停止功能，服务将在指定时间后自动停止，不再产生费用。

1.7 使用窍门

1.7.1 创建项目时，如何快速创建 OBS 桶及文件夹？

在创建项目时需要选择训练数据路径，本章节将指导您如何在选择训练数据路径时，快速创建OBS桶和OBS文件夹。

1. 在创建自动学习项目页面，单击数据集输入位置右侧的“📁”按钮，进入“数据集输入位置”对话框。
2. 单击“新建对象存储服务（OBS）桶”，进入创建桶页面，具体请参见《对象存储服务控制台指南》中的[创建桶](#)章节。

图 1-45 快速创建 OBS 桶



3. 桶创建完成后，选择对应桶名称，单击“新建文件夹”，在“新建文件夹”对话框中，填写文件夹“名称”，单击“确定”完成创建，选择创建的文件夹。
 - 文件夹名称不能包含以下字符：\:*? "<>|。
 - 文件夹名称不能以英文句号（.）或斜杠（/）开头或结尾。
 - 文件夹的绝对路径总长度不能超过1023字符。
 - 任何单个斜杠（/）表示分隔并创建多层级的文件夹。

图 1-46 新建文件夹

数据集输入位置



1.7.2 自动学习生成的模型，存储在哪里？支持哪些其他操作？

模型统一管理

针对自动学习项目，当模型训练完成后，其生成的模型，将自动进入“AI应用管理 > AI应用”页面，如下图所示。模型名称由系统自动命名，前缀与自动学习项目的名称一致，方便辨识。

注意

自动学习生成的模型，不支持下载使用。

图 1-47 自动学习生成的模型

AI应用名称	最新版本	状态	部署类型
huashuinih@ip-xxx	0.0.1	正常	在线服务
MA-Service-xxxx-11-15-18-54	1.0.0	正常	在线服务/批量服务/边缘服务
model-0x44T185-flower	0.0.1	正常	在线服务/批量服务/边缘服务

自动学习生成的模型，支持哪些其他操作

- 支持部署为在线服务、批量服务或边缘服务。**
 在自动学习页面中，仅支持部署为在线服务，如需部署为批量服务或边缘服务，可在“AI应用管理 > AI应用”页面中直接部署。
- 支持发布至市场**
 将产生的模型发布至AI Gallery，共享给其他用户。
- 支持创建新版本**
 创建新版本，仅支持从ModelArts训练作业、OBS、模型模板、或自定义镜像中选择元模型。无法从原自动学习项目中，创建新版本。
- 支持删除模型或其模型版本**

2 自动学习（旧版）

2.1 自动学习简介

自动学习功能介绍

ModelArts自动学习是帮助人们实现AI应用的低门槛、高灵活、零代码的定制化模型开发工具。自动学习功能根据标注数据自动设计模型、自动调参、自动训练、自动压缩和部署模型。开发者无需专业的开发基础和编码能力，只需上传数据，通过自动学习界面引导和简单操作即可完成模型训练和部署。

当前自动学习支持快速创建图像分类、物体检测、预测分析、声音分类和文本分类模型的定制化开发。可广泛应用在工业、零售安防等领域。

- 图像分类：识别图片中物体的类别。
- 物体检测：识别出图片中每个物体的位置和类别。
- 预测分析：对结构化数据做出分类或数值预测。
- 声音分类：对环境中不同声音进行分类识别。
- 文本分类：识别一段文本的类别。文本分类目前只支持中文。

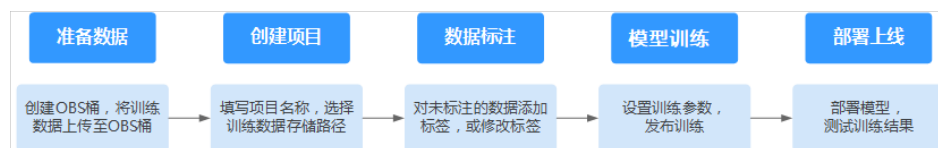
📖 说明

旧版自动学习仅支持使用旧版数据集功能，不支持使用新版数据集功能。

自动学习流程介绍

使用ModelArts自动学习开发AI模型无需编写代码，您只需上传数据、创建项目、完成数据标注、发布训练、然后将训练的模型部署上线。您最多只能创建100个自动学习项目。具体流程请参见图2-1。

图 2-1 自动学习操作流程



项目类型介绍

- **图像分类**

图像分类项目，是对图像进行分类。添加图片并对图像进行分类标注。完成图片标注后开始模型训练，即可快速生成图像分类模型。可应用于商品的自动分类、运输车辆种类识别和残次品的自动分类。例如质量检查的场景，则可以上传产品图片，将图片标注“合格”、“不合格”，通过训练部署模型，实现产品的质检。

- **物体检测**

物体检测项目，是检测图片中物体的类别与位置。需要添加图片，用合适的框标注物体作为训练集，进行训练输出模型。适用于一张图片中要识别多个物体或者物体的计数等。可应用于园区人员穿戴规范检测和物品摆放的无人巡检。

- **预测分析**

预测分析项目，是一种针对结构化数据的模型自动训练应用，能够对结构化数据进行分类或者数据预测。可用于用户画像分析，实现精准营销。也可应用于制造设备预测性维护，根据设备实时数据的分析，进行故障识别。

- **声音分类**

声音分类项目，是识别一段音频中是否包含某种声音。可应用于生产或安防场景的异常声音监控。

- **文本分类**

文本分类项目，识别一段文本的类别。

2.2 图像分类

2.2.1 准备数据

使用ModelArts自动学习构建模型时，您需要将数据上传至对象存储服务（OBS）中。OBS桶需要与ModelArts在同一区域。

数据上传至 OBS

在本文档中，采用管理控制台上传数据至OBS。

执行如下操作，将数据导入到您的数据集中，以便用于模型训练和构建。

1. 登录OBS管理控制台，在ModelArts同一区域内**创建桶**。如果已存在可用的桶，需确保OBS桶与ModelArts在同一区域。
2. 参考**上传文件**，将本地数据上传至OBS桶中。如果您的数据较多，推荐OBS Browser+上传数据或上传文件夹。上传的数据需满足此类型自动学习项目的数据集要求。

说明

在上传数据时，请选择非加密桶进行上传，否则会由于加密桶无法解密导致后期的训练失败。

数据集要求

- 文件名规范:不能有+、空格、制表符。
- 保证图片质量:不能有损坏的图片，目前支持的格式包括jpg、jpeg、bmp、png。

- 不要把明显不同的多个任务数据放在同一个数据集内。
- 每一类数据尽量多，尽量均衡。期望获得良好效果，图像分类项目中，建议训练数据集保证每类图片超过100张。
- 为了保证模型的预测准确度，训练样本跟真实使用场景尽量相似。
- 为保证模型的泛化能力，数据集尽量覆盖可能出现的各种场景。

OBS 上传文件规范

- 如不需要提前上传训练数据，请创建一个空文件夹用于存放工程后期生成的文件。如：“/bucketName/data-cat”。
- 如需要提前上传待标注的图片，请创建一个空文件夹，然后将图片文件保存在该文件夹下，图片的目录结构如：“/bucketName/data-cat/cat.jpg”。
- 如您将已标注好的图片上传至OBS桶，请按照如下规范上传。
 - 图像分类数据集要求将标注对象和标注文件存储在同一目录，并且一一对应，例如标注对象文件名为“10.jpg”，那么标注文件的文件名应为“10.txt”。

数据文件存储示例：

```
|-<dataset-import-path>
|
|   10.jpg
|   10.txt
|   11.jpg
|   11.txt
|   12.jpg
|   12.txt
```

- 只支持JPG、JPEG、PNG、BMP格式的图片。在OBS管理控制台上传时，单张图片的大小不能超过5MB，单次上传的图片总大小不能超过8MB，数据量大时推荐使用OBS Browser+上传。
- 标签名是由中文、大小写字母、数字、中划线或下划线组成，且不超过32位的字符串。
- 图像分类标签“.txt”规范如下。
 - 一行一个标签：

```
flower
book
...
```

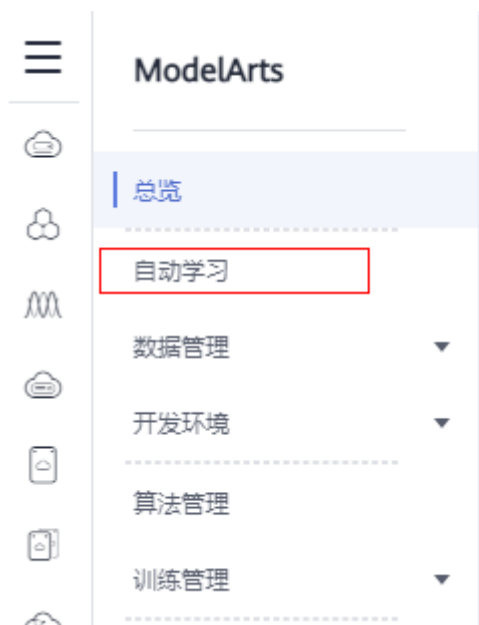
2.2.2 创建项目

ModelArts自动学习，包括图像分类项目、物体检测项目、预测分析项目、声音分类和文本分类项目。您可以根据业务需求选择创建合适的项目。您需要执行如下操作来创建自动学习项目。

创建项目

1. 登录ModelArts管理控制台，在左侧导航栏单击“自动学习”，单击“返回旧版”进入旧版“自动学习”页面。

图 2-2 进入自动学习



2. 在您需要的自动学习项目列表中，单击“创建项目”，进入创建自动学习项目界面。

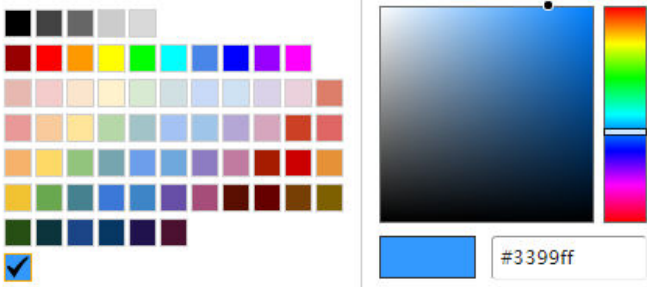
图 2-3 自动学习列表



3. 在创建自动学习项目页面，计费模式默认“按需计费”，参考表2-1填写相应参数。

表 2-1 参数说明

参数	说明
“名称”	项目的名称。 <ul style="list-style-type: none"> 名称只能包含数字、字母、下划线和中划线，长度不能超过32位且不能为空。 名称请以大写或小写字母开头。
“描述”	对项目的简要描述。

参数	说明
“数据集来源”	<p>可选择“新建数据集”或“已有数据集”。</p> <ul style="list-style-type: none"> “新建数据集”：需填写创建数据集所需的参数，包含“数据集名称”、“数据集输入位置”、“数据集输出位置”和“添加标签集”信息。 “已有数据集”：您可以从ModelArts数据管理中选择一个同类型的数据集，用于创建自动学习项目。“数据集名称”的下拉框中，仅展示同类型的数据集供选择。
“数据集名称”	<p>当选择“新建数据集”时，此参数右侧为文本框，需根据规则填写新创建的数据集名称；当选择“已有数据集”时，此参数右侧为下拉框，展示当前账号下可用的同类型数据集列表。</p>
“数据集输入位置”	<p>选择数据集输入位置的OBS路径。数据集输入的规范请参见准备数据章节。</p> <ul style="list-style-type: none"> 除“准备数据>OBS上传文件规范”说明的文件及文件夹外，训练数据路径下不允许存任何其他文件或文件夹，否则会报错。 训练数据路径下的文件请不要随意修改。
“数据集输出位置”	<p>选择数据集输出位置的OBS路径。</p> <p>说明 “数据集输出位置”不能与“数据集输入位置”为同一路径，且不能是“数据集输入位置”的子目录。“数据集输出位置”最好选择一个空目录。</p>
“添加标签集”	<ul style="list-style-type: none"> 设置标签名称：在标签名称文本框中，输入标签名称。标签名称只能是中文、字母、数字、下划线或中划线组成的合法字符串。长度为1~32字符。 添加标签：单击“添加标签”可增加多个标签。 设置标签颜色：“物体检测”和“文本分类”的数据集需设置标签颜色，“图像分类”和“声音分类”类型不需要。在每个标签右侧的标签颜色区域下，可在色板中选择颜色，或者直接输入十六进制颜色码进行设置。 

- 单击“创建项目”，待页面提示“创建项目成功”完成创建，创建成功后页面自动跳转到“数据标注”。您也可以直接在“自动学习”页面查看创建成功的项目，单击项目名称进入“数据标注”。

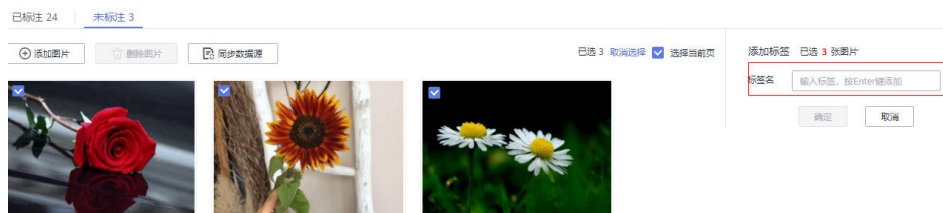
2.2.3 数据标注

由于模型训练过程需要大量有标签的图片数据，因此在模型训练之前需对没有标签的图片添加标签。通过ModelArts您可对图片进行一键式批量添加标签，快速完成对图片的标注操作，也可以对已标注图片修改或删除标签进行重新标注。用于训练的图片，至少有2种以上的分类，每种分类的图片数不少于5张，建议至少50张图片，则效果会更好，若图像分类相似度较高，则需要更多的图片。

图片标注

1. 在“数据标注”页面单击“未标注”页签，此页面展示所有未标注的图片数据。依次单击选中待标注的图片，或勾选“选择当前页”选中该页面所有图片，在页面右侧进行图片标注。

图 2-4 图片标注



2. 选中图片后，在页面右侧“添加标签”，输入“标签名”或从下拉列表中选择已添加的标签。单击“确定”，完成选中图片的标注操作。例如，您可以选择多张图片，按照花朵种类将图片标注为“tulips”。同样选择其他未标注分类图片，将其标注为“sunflowers”、“roses”等。标注完成后，图片将存储至“已标注”页签下。
 - a. 图片标注支持多标签，即一张图片可添加多个标签。
 - b. 标签名是由中文、大小写字母、数字、中划线或下划线组成，且不超过32位的字符串。

图 2-5 图像分类图片标注



3. 当图片目录中所有图片都完成标注后，您可以在“已标注”页签下查看已完成标注的图片，或者通过右侧的“全部标签”列表，了解当前已完成的标签名称和标签数量。

同步或添加图片

在“自动学习”页面，单击项目名称，进入“自动学习 > 数据标注”页面。项目创建时，数据标注的图片来源有两种，通过本地添加图片和同步OBS中的图片数据。

图 2-6 添加本地图片



图 2-7 同步 OBS 图片数据



- **添加图片**：您可以将本地图片快速添加到ModelArts，同时自动上传至创建项目时所选择的OBS路径中。单击“添加图片”，在弹出的对话框中单击“添加图片”并添加。一次上传所有图片的总大小不能超过8MB。单张图片大小不能超过5MB。
- **同步数据源**：将图片数据上传至创建项目时指定的OBS目录，然后单击“同步数据源”，快速将通过OBS目录中的图片数据添加到ModelArts。
- **删除图片**：您可以依次单击选中图片进行删除，也可以勾选“选择当前页”对该页面所有图片进行删除。

📖 说明

所有的删除操作均不可恢复，请谨慎操作。

修改标注

当数据完成标注后，您还可以进入已标注页签，对已标注的数据进行修改。

• 基于图片修改

在数据标注页面，单击“已标注”页签，然后在图片列表中选中待修改的图片（选择一个或多个）。在右侧标签信息区域中对图片信息进行修改。


- **添加标签**：在“标签名”右侧文本框中，选择已有标签或输入新的标签名，然后单击“确定”，为选中图片增加标签。
- **修改标签**：在“选中文件标签”区域中，单击操作列的编辑图标，然后在文本框中输入正确的标签名，然后单击  图标完成修改。

图 2-8 编辑标签

选中文件标签		
标签名称	标签数量	操作
其他垃圾_烟蒂	1	 


- 删除标签：在“选中文件标签”区域中，单击操作列的删除该标签。
- **基于标签修改**
在数据标注页面，单击“已标注”页签，在图片列表右侧，显示全部标签的信息。

图 2-9 全部标签的信息

全部标签 8		
标签名称	标签数量 	操作
_烟蒂	100	 
厨余垃圾_蛋壳	100	 
厨余垃圾_水...	100	 

- 修改标签：单击操作列的编辑按钮，然后在弹出的对话框中输入修改后的标签名，然后单击“确定”完成修改。修改后，之前添加了此标签的图片，都将被标注为新的标签名称。
- 删除标签：单击操作列的删除按钮，在弹出的对话框中，选择“仅删除标签”、“删除标签及仅包含此标签的图片（不删除源文件）”或“删除标签及仅包含此标签的图片（同时删除源文件）”，然后单击“确定”。

图 2-10 删除标签

删除



确定要删除daisy标签吗？

此标签下有7张图片，删除操作无法恢复，请谨慎操作。

- 仅删除标签
- 删除标签及仅包含此标签的图片（不删除源文件）
- 删除标签及仅包含此标签的图片（同时删除源文件） 

确定

取消

2.2.4 模型训练

完成图片标注后，可进行模型的训练。模型训练的目的在于得到满足需求的图像分类模型。由于用于训练的图片，至少有2种以上的分类，每种分类的图片数不少于5张。因此在发布训练之前，请确保已标注的图片符合要求，否则右上方“开始训练”按钮会处于灰色状态。

操作步骤

1. 在“自动学习”页面，单击创建成功的项目名称，进入“数据标注”页面，完成数据标注。

图 2-11 完成数据标注



2. 在“数据标注”页面，单击右上角的“开始训练”，然后在弹出的“训练设置”对话框中配置相关参数。参数说明请参见表2-2。请确保“训练验证比例”输入值的小数位数应在1~5之间。

表 2-2 训练设置参数说明

参数	说明	默认值
数据集版本名称	此版本即数据管理中发布数据集时设置的版本。自动学习项目中，启动训练作业时，会基于前面的数据标注，将数据集发布为一个版本。 系统将自动给出一个版本号，您也可以根据实际情况进行填写。	系统随机给出
训练验证比例	训练验证比例表示将已标注样本随机分为训练集和验证集的比例，默认训练集比例为0.8，即训练集占0.8，验证集占0.2。manifest中的usage字段记录划分类别。取值范围为0~1。	0.8
增量训练版本	用户可以在之前训练成功的版本中，自主选择精度最高的版本进行再训练，可以加快模型收敛速度，提高训练精度。	无
最大训练时长（分钟）	即最大训练时长，在该时长内若训练还未完成，则保存模型停止训练。为防止模型未收敛就退出，建议使用较大值。输入值取值范围为6~6000。建议适当延长训练时间。	60

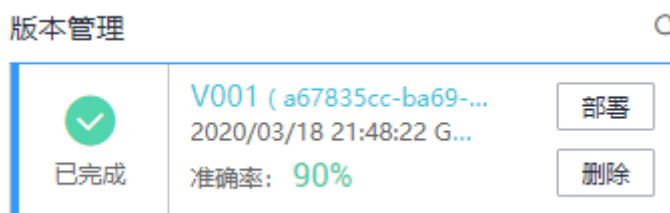
参数	说明	默认值
训练偏好	<ul style="list-style-type: none"> performance_first: 性能优先, 训练时间较短, 模型较小 balance: 平衡 accuracy_first: 精度优先, 训练时间较长, 模型较大 	balance
计算规格	选择训练使用的资源规格。	以控制台提供为准。

3. 训练参数设置完成后, 单击“下一步”进入配置页, 确认规格后单击“提交”进行模型的自动训练, 训练时间相对较长, 建议您耐心等待。如果关闭或退出此页面, 系统仍然在执行训练操作。

如果使用免费规格, 还需仔细阅读界面提示, 同时勾选“我已阅读并同意以上内容”。

4. 在“模型训练”页签中, 待训练状态由“运行中”变为“已完成”, 即完成模型的自动训练。

图 2-12 运行成功



5. 训练完成后, 您可以在界面中查看训练详情, 如“准确率”、“评估结果”、“训练参数”、“分类统计表”等。评估结果参数说明请参见表2-3。

图 2-13 模型训练结果

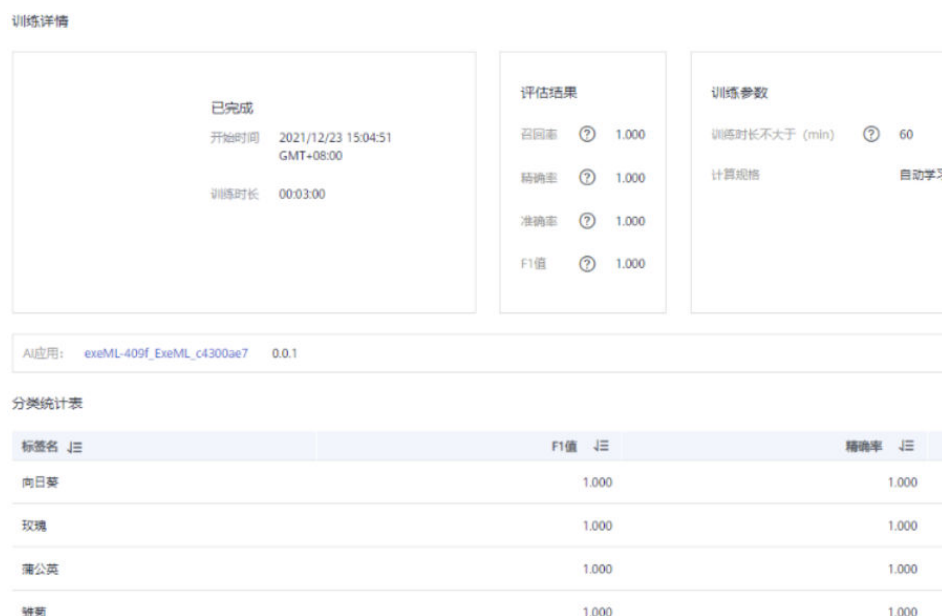


表 2-3 评估结果参数说明

参数	说明
召回率	被用户标注为某个分类的所有样本中，模型正确预测为该分类的样本比率，反映模型对正样本的识别能力。
精确率	被模型预测为某个分类的所有样本中，模型正确预测的样本比率，反映模型对负样本的区分能力。
准确率	所有样本中，模型正确预测的样本比率，反映模型对样本整体的识别能力。
F1值	F1值是模型精确率和召回率的加权调和平均，用于评价模型的好坏，当F1较高时说明模型效果较好。

说明

同一个自动学习项目可以训练多次，每次训练生成一个版本。如第一次训练版本号为“V001 (xxx)”，下一个版本为“V002 (xxx)”。基于训练版本可以对训练模型进行管理。当训练的模型达到目标后，再执行部署上线的操作。

2.2.5 部署上线

部署上线

部署上线操作即将模型部署为在线服务，并且提供在线的测试UI与监控能力。完成模型训练后，可选择准确率理想且训练状态为“运行成功”的版本部署上线。具体操作步骤如下。

1. 在“模型训练”页签中，待训练状态变为“运行成功”，单击版本管理区域中的“部署”，开始将模型部署上线为在线服务。

图 2-14 部署操作



2. 在弹出的“部署”对话框中，选择资源规格，同时设置自动停止功能，然后单击确定，启动部署。
 - “计算节点规格”：以控制台实际提供为准。
 - “计算节点个数”：默认为1，且不能修改。
 - “是否自动停止”：启用该参数并设置时间后，服务将在指定时间后自动停止。如果不启用此参数，在线服务将一直运行，同时一直收费，自动停止功能可以帮您避免产生不必要的费用。默认开启自动停止功能，且默认值为“1小时后”。

目前支持设置为“1小时后”、“2小时后”、“4小时后”、“6小时后”、“自定义”。如果选择“自定义”的模式，可在右侧输入框中输入1~24范围内的任意整数。

3. 启动部署上线后，可以在“部署上线”界面查看模型部署上线的状态。部署上线将耗费较多时间，请您耐心等待。当“部署上线”页签版本管理区域的状态由“部署中”变更为“运行中”，部署完成。

📖 说明

在自动学习界面中，仅支持将训练后的模型部署为在线服务，如果需要部署为“批量服务”或“边缘服务”，请参见[自动学习生成的模型，存储在哪里？支持哪些其他操作？](#)。

服务测试

- 您可以在“部署上线”页面，选择对应的服务类型，例如自动学习图像分类项目默认将服务部署为在线服务，进入“在线服务”页面，单击目标服务“操作”列的“预测”，进行服务测试，测试方法和下方陈述操作步骤一致。具体操作请参见[测试服务](#)。
- 您也可以通过调用代码对服务进行测试，根据部署服务类型的不同，具体操作详情参见[访问在线服务](#)、[访问边缘服务](#)。
- 下面的测试，是您在自动学习图像分类项目页面将模型部署上线之后进行服务测试的操作步骤。
 - a. 模型部署完成后，您可添加图片进行测试。在“自动学习”页面，选择目标项目，进入“部署上线”界面，选择状态为“运行中”的服务版本，在“服务测试”区域单击“上传”，选择本地图片进行测试。

图 2-15 上传图片

服务测试

当服务状态为“ 运行中”时，才能进行预测。

选择需要预测的文件

上传

预测

- b. 单击“预测”进行测试，预测完成后，右侧“预测结果”区域输出标签名称“sunflowers”和检测的评分。如模型准确率不满足预期，可在“数据标注”页签中添加图片并进行标注，重新进行模型训练及部署上线。预测结果中的参数说明请参见[表2-4](#)。如果您对模型预测结果满意，可根据界面提示调用接口访问在线服务，操作指导请参见[“访问在线服务”](#)。
目前只支持jpg、jpeg、bmp、png格式的图片。

图 2-16 预测结果

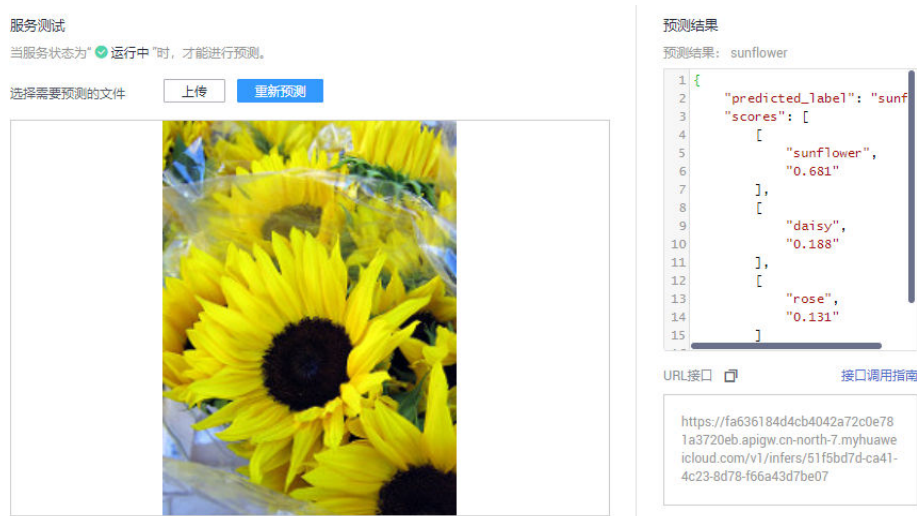


表 2-4 预测结果中的参数说明

参数	说明
predict_label	表示图片预测的标签。
scores	表示Top5标签的预测置信度。

说明

由于“运行中”的在线服务将持续耗费资源，如果不再使用此在线服务，建议在版本管理区域，单击“停止”，即可停止在线服务的部署，避免产生不必要的费用。如果需要继续使用此服务，可单击“启动”恢复。

如果您启用了自动停止功能，服务将在指定时间后自动停止，不再产生费用。

2.3 物体检测

2.3.1 准备数据

使用ModelArts自动学习构建模型时，您需要将数据上传至对象存储服务（OBS）中。OBS桶需要与ModelArts在同一区域。

数据上传至 OBS

在本文档中，采用管理控制台上传数据至OBS。

执行如下操作，将数据导入到您的数据集中，以便用于模型训练和构建。

1. 登录OBS管理控制台，在ModelArts同一区域内**创建桶**。如果已存在可用的桶，需确保OBS桶与ModelArts在同一区域。
2. 参考**上传文件**，将本地数据上传至OBS桶中。如果您的数据较多，推荐OBS Browser+上传数据或上传文件夹。上传的数据需满足此类型自动学习项目的数据集要求。

📖 说明

在上传数据时，请选择非加密桶进行上传，否则会由于加密桶无法解密导致后期的训练失败。

数据集要求

- 文件名规范，不能有中文，不能有+、空格、制表符。
- 保证图片质量：不能有损坏的图片；目前支持的格式包括jpg、jpeg、bmp、png。
- 不要把明显不同的多个任务数据放在同一个数据集内。
- 为了保证模型的预测准确度，训练样本跟真实使用场景尽量相似。
- 为保证模型的泛化能力，数据集尽量覆盖可能出现的各种场景。
- 物体检测数据集中，如果标注框坐标超过图片，将无法识别该图片为已标注图片。

OBS 上传文件规范

- 如不需要提前上传训练数据，请创建一个空文件夹用于存放工程后期生成的文件。如：“/bucketName/data-cat”。
- 如需要提前上传待标注的图片，请创建一个空文件夹，然后将图片文件保存在该文件夹下，图片的目录结构如：“/bucketName/data-cat/cat.jpg”。
- 如您将已标注好的图片上传至OBS桶，请按照如下规范上传。
 - 物体检测数据集要求用户将标注对象和标注文件存储在同一目录，并且一一对应。例如标注对象文件名为“IMG_20180919_114745.jpg”，那么标注文件的文件名应为“IMG_20180919_114745.xml”。

物体检测的标注文件需要满足PASCAL VOC格式，格式详细说明请参见 [表 2-5](#)。

数据存储示例：

```

|<dataset-import-path>
|   IMG_20180919_114732.jpg
|   IMG_20180919_114732.xml
|   IMG_20180919_114745.jpg
|   IMG_20180919_114745.xml
|   IMG_20180919_114945.jpg
|   IMG_20180919_114945.xml
    
```

- 只支持JPG、JPEG、PNG、BMP格式的图片，在OBS管理控制台上传时，单张图片的大小不能超过5MB，单次上传的图片总大小不能超过8MB，数据量大时推荐使用OBS Browser+上传。
- 标签名是由中文、大小写字母、数字、中划线或下划线组成，且不超过32位的字符串。

表 2-5 PASCAL VOC 格式说明

字段	是否必选	说明
folder	是	表示数据源所在目录。
filename	是	被标注文件的文件名。

字段	是否必选	说明
size	是	表示图像的像素信息。 <ul style="list-style-type: none"> width: 必选字段，图片的宽度。 height: 必选字段，图片的高度。 depth: 必选字段，图片的通道数。
segmented	是	表示是否用于分割。
object	是	表示物体检测信息，多个物体标注会有多个object体。 <ul style="list-style-type: none"> name: 必选字段，标注内容的类别。 pose: 必选字段，标注内容的拍摄角度。 truncated: 必选字段，标注内容是否被截断（0表示完整）。 occluded: 必选字段，标注内容是否被遮挡（0表示未遮挡）。 difficult: 必选字段，标注目标是否难以识别（0表示容易识别）。 confidence: 可选字段，标注目标的置信度，取值范围0-1之间。 bndbox: 必选字段，标注框的类型，标注信息请参见 表2-6。

表 2-6 标注框类型描述

type	形状	标注信息
bndbox	矩形框	左上和右下两个点坐标。 <xmin>100<xmin> <ymin>100<ymin> <xmax>200<xmax> <ymin>200<ymin>

标注文件示例：

```
<annotation>
  <folder>test_data</folder>
  <filename>260730932.jpg</filename>
  <size>
    <width>767</width>
    <height>959</height>
    <depth>3</depth>
  </size>
  <segmented>0</segmented>
  <object>
    <name>bag</name>
    <pose>Unspecified</pose>
```

```
<truncated>0</truncated>  
<occluded>0</occluded>  
<difficult>0</difficult>  
<bndbox>  
  <xmin>108</xmin>  
  <ymin>101</ymin>  
  <xmax>251</xmax>  
  <ymax>238</ymax>  
</bndbox>  
</object>  
</annotation>
```

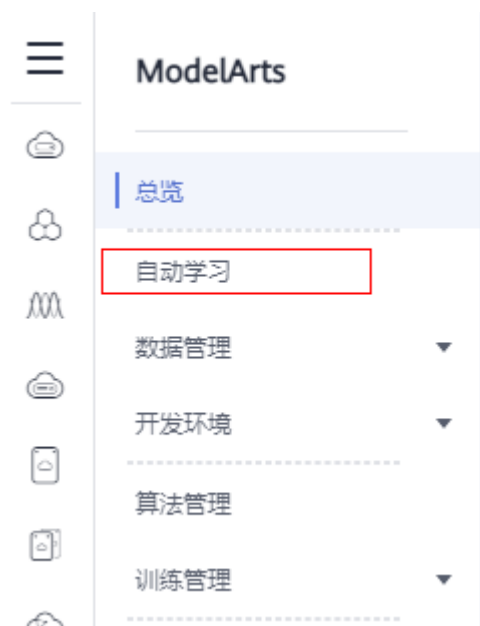
2.3.2 创建项目

ModelArts自动学习，包括图像分类项目、物体检测项目、预测分析项目、声音分类和文本分类项目。您可以根据业务需求选择创建合适的项目。您需要执行如下操作来创建自动学习项目。

创建项目

1. 登录ModelArts管理控制台，在左侧导航栏单击“自动学习”，单击“返回旧版”进入旧版“自动学习”页面。

图 2-17 进入自动学习



2. 在您需要的自动学习项目列表中，单击“创建项目”，进入创建自动学习项目界面。

图 2-18 自动学习列表



3. 在创建自动学习项目页面，计费模式默认“按需计费”，参考表2-7填写相应参数。

表 2-7 参数说明

参数	说明
“名称”	项目的名称。 <ul style="list-style-type: none">名称只能包含数字、字母、下划线和中划线，长度不能超过32位且不能为空。名称请以大写或小写字母开头。
“描述”	对项目的简要描述。
“数据集来源”	可选择“新建数据集”或“已有数据集”。 <ul style="list-style-type: none">“新建数据集”：需填写创建数据集所需的参数，包含“数据集名称”、“数据集输入位置”、“数据集输出位置”和“添加标签集”信息。“已有数据集”：您可以从ModelArts数据管理中选择一个同类型的数据集，用于创建自动学习项目。“数据集名称”的下拉框中，仅展示同类型的数据集供选择。
“数据集名称”	当选择“新建数据集”时，此参数右侧为文本框，需根据规则填写新创建的数据集名称；当选择“已有数据集”时，此参数右侧为下拉框，展示当前账号下可用的同类型数据集列表。
“数据集输入位置”	选择数据集输入位置的OBS路径。数据集输入的规范请参见“数据准备”章节。 <ul style="list-style-type: none">除“准备数据>OBS上传文件规范”说明的文件及文件夹外，训练数据路径下不允许存任何其他文件或文件夹，否则会报错。训练数据路径下的文件请不要随意修改。
“数据集输出位置”	选择数据集输出位置的OBS路径。 说明 “数据集输出位置”不能与“数据集输入位置”为同一路径，且不能是“数据集输入位置”的子目录。“数据集输出位置”最好选择一个空目录。

参数	说明
“添加标签集”	<ul style="list-style-type: none"> ● 设置标签名称：在标签名称文本框中，输入标签名称。标签名称只能是中文、字母、数字、下划线或中划线组成的合法字符串。长度为1~32字符。 ● 添加标签：单击“添加标签”可增加多个标签。 ● 设置标签颜色：“物体检测”和“文本分类”的数据集需设置标签颜色，“图像分类”和“声音分类”类型不需要。在每个标签右侧的标签颜色区域下，可在色板中选择颜色，或者直接输入十六进制颜色码进行设置。 

4. 单击“创建项目”，待页面提示“创建项目成功”完成创建，创建成功后页面自动跳转到“数据标注”。您也可以直接在“自动学习”页面查看创建成功的项目，单击项目名称进入“数据标注”。

2.3.3 数据标注

物体检测之前，首先需考虑如何设计标签，标签设计需要对应所检测图片的明显特征，并且选择的标签比较容易识别（画面主体物与背景区分度较高），每个标签就是对所检测图片期望识别的全部结果。物体的标签设计完成之后，基于设计好的标签准备该图片的数据，每种需识别出的标签，建议应在所有图片个数相加超过100张，若某些图片的标签具有相似性，则需要更多的图片。

- 标注时，类内方差尽量要小。即相同类别的标注，尽量近似；不同类别的标注，尽量保持差距较大。
- 标记的每个标签尽量和背景有较大的区分度。
- 物体检测标注，需要保证目标框内物体的完整性；针对图片中存在多个物体的情形，做到不重标、不漏标。

图片标注

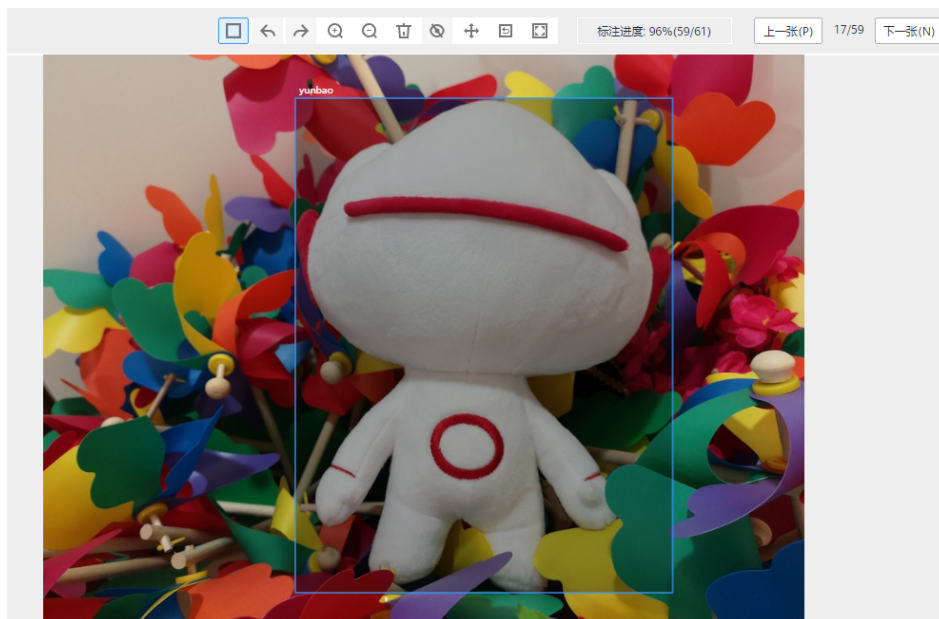
1. 在“数据标注”页面单击“未标注”页签，此页面展示所有未标注的图片数据。单击任意一张图片，进入图片标注界面。
2. 用鼠标框选图片中的物体所在区域，然后在弹出的对话框中选择标签颜色，输入标签名称，例如此示例中的“yunbao”，按“Enter”键完成此标签的添加。标注完成后，左侧图片目录中此图片的状态将显示为“已标注”。

数据标注的更多说明：

- 您可以在图片上方或下方单击左右切换键，或者按键盘的左右方向键，选择其他图片，重复上述操作继续进行图片标注。如果一张图片有多个物体，您可以标注多处。

- 同一个物体检测自动学习项目内，可以增加多个标签，且标签可选择不同颜色，方便识别。使用鼠标完成物体框选后，在弹出的对话框中，选择新的颜色，输入新的标签名称，即可添加一个新的标签。
- 自动学习项目中，物体检测仅支持矩形标注框。在“数据管理”功能中，物体检测类型的数据集，支持更多类型的标注框。
- 在标注窗口中，您可以滚动鼠标，放大或缩小图片，方便您快速定位到物体位置。

图 2-19 物体检测图片标注



3. 当图片目录中所有图片都完成标注后，单击左上角“自动学习”，在弹出的对话框中单击“确定”保存标注信息。页面将进入数据标注页面，可以在“已标注”页签下查看已完成标注的图片，或者通过右侧的标签信息，了解当前已完成的标签名称和标签数量。

同步或添加图片

在“自动学习”页面，单击项目名称，进入“自动学习 > 数据标注”页面。项目创建时，数据标注的图片来源有两种，通过本地添加图片和同步OBS中的图片数据。

图 2-20 添加本地图片



图 2-21 同步 OBS 图片数据



- **添加图片**：您可以将本地图片快速添加到ModelArts，同时自动上传至创建项目时所选择的OBS路径中。单击“添加图片”，在弹出的对话框中单击“添加图片”并添加。一次上传所有图片的总大小不能超过8MB。单张图片大小不能超过5MB。
- **同步数据源**：将图片数据上传至创建项目时指定的OBS目录，然后单击“同步数据源”，快速将通过OBS目录中的图片数据添加到ModelArts。
- **删除图片**：您可以依次单击选中图片进行删除，也可以勾选“选择当前页”对该页面所有图片进行删除。

📖 说明

所有的删除操作均不可恢复，请谨慎操作。

修改标注

当数据完成标注后，您还可以进入已标注页签，对已标注的数据进行修改。

● 基于图片修改

在数据集详情页面，单击“已标注”页签，然后在图片列表中选中待修改的图片，单击该图片跳转到标注页面，在右侧“标注”信息区域中对图片信息进行修改。


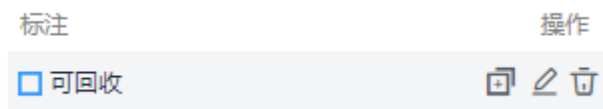
- 修改标签：“标注”区域中，单击编辑按钮，在文本框中输入正确的标签名，然后单击  按钮完成修改。标签颜色不支持修改。
- 删除标签：在“标注”区域中，单击删除按钮，即可删除此图片中的标签。标签删除后，单击页面左上角的项目名称离开标注页面。该图片会重新回到“未标注”页签。

图 2-22 编辑物体检测标签



● 基于标签修改

在数据集详情页面，单击“已标注”页签，在图片列表右侧，显示全部标签的信息。单击操作列的编辑按钮，然后在弹出的对话框中输入修改后的标签名，然后单击“确定”完成修改。修改后，之前添加了此标签的图片，都将被标注为新的标签名称。

图 2-23 物体检测的全部标签

全部标签 2

标签名称	标签数量	样本数量	操作
厨余...	0	0	
可回收	1	1	

2.3.4 模型训练

自动学习物体检测项目，在图片标注完成后，通过模型训练得到合适的模型版本。

操作步骤

1. 在“自动学习”页面，单击创建成功的项目名称，进入“数据标注”页面，完成数据标注。

图 2-24 完成数据标注



2. 在“数据标注”页面，单击右上角的“开始训练”，然后在弹出的“训练设置”对话框中配置相关参数。参数说明请参见表2-8。请确保训练验证比例输入值的小数位应在1~5之间。

表 2-8 训练设置参数说明

参数	说明	默认值
数据集版本名称	此版本即数据管理中发布数据集时设置的版本。自动学习项目中，启动训练作业时，会基于前面的数据标注，将数据集发布为一个版本。 系统将自动给出一个版本号，您也可以根据实际情况进行填写。	系统随机给出
训练验证比例	训练验证比例表示将已标注样本随机分为训练集和验证集的比例，默认训练集比例为0.8，即训练集占0.8，验证集占0.2。manifest中的usage字段记录划分类别。取值范围为0~1。	0.8

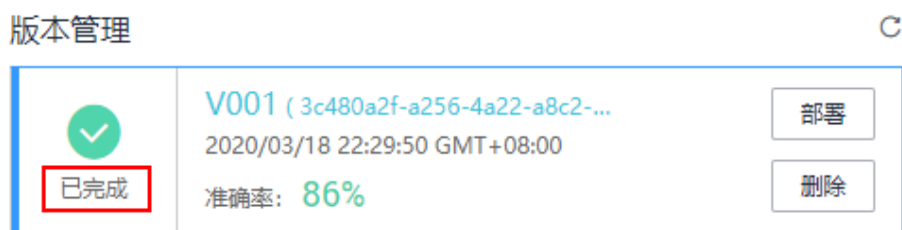
参数	说明	默认值
增量训练版本	用户可以在之前训练成功的版本中，自主选择精度最高的版本进行再训练，可以加快模型收敛速度，提高训练精度。	无
最大训练时长（分钟）	即最大训练时长，在该时长内若训练还未完成，则保存模型停止训练。为防止模型未收敛就退出，建议使用较大值。输入值取值范围为6~6000。建议适当延长训练时间，2000张图片的训练集建议选择运行1小时以上。	60
训练偏好	<ul style="list-style-type: none"> performance_first: 性能优先，训练时间较短，模型较小 balance: 平衡 accuracy_first: 精度优先，训练时间较长，模型较大 	balance
计算规格	选择训练使用的资源规格。	以控制台提供为准。

- 训练参数设置完成后，单击“下一步”进入配置页，确认规格后单击“提交”进行模型的自动训练，训练时间相对较长，建议您耐心等待。如果关闭或退出此页面，系统仍然在执行训练操作。

如果使用免费规格，还需仔细阅读界面提示，同时勾选“我已阅读并同意以上内容”。

- 在“模型训练”页签中，待训练状态由“运行中”变为“已完成”，即完成模型的自动训练。

图 2-25 运行成功



- 训练完成后，您可以在界面中查看训练详情，如“准确率”、“评估结果”、“训练参数”、“分类统计表”等。评估结果参数说明请参见表2-9。

图 2-26 模型训练结果



表 2-9 评估结果参数说明

参数	说明
召回率	被用户标注为某个分类的所有样本中，模型正确预测为该分类的样本比率，反映模型对正样本的识别能力。
精确率	被模型预测为某个分类的所有样本中，模型正确预测的样本比率，反映模型对负样本的区分能力。
准确率	所有样本中，模型正确预测的样本比率，反映模型对样本整体的识别能力。
F1值	F1值是模型精确率和召回率的加权调和平均，用于评价模型的好坏，当F1较高时说明模型效果较好。

📖 说明

同一个自动学习项目可以训练多次，每次训练生成一个版本。如第一次训练版本号为“V001 (xxx)”，下一个版本为“V002 (xxx)”。基于训练版本可以对训练模型进行管理。当训练的模型达到目标后，再执行部署上线的操作。

2.3.5 部署上线

部署上线

部署上线操作即将模型部署为在线服务，并且提供在线的测试UI与监控能力。完成模型训练后，可选择准确率理想且训练状态为“运行成功”的版本部署上线。具体操作步骤如下。

1. 在“模型训练”页签中，待训练状态变为“运行成功”，单击版本管理区域中的“部署”，开始将模型部署上线为在线服务。

图 2-27 部署操作



2. 在弹出的“部署”对话框中，选择资源规格，同时设置自动停止功能，然后单击确定，启动部署。
 - “计算节点规格”：以控制台实际提供为准。
 - “计算节点个数”：默认为1，且不能修改。
 - “是否自动停止”：启用该参数并设置时间后，服务将在指定时间后自动停止。如果不启用此参数，在线服务将一直运行，同时一直收费，自动停止功能可以帮您避免产生不必要的费用。默认开启自动停止功能，且默认值为“1小时后”。

目前支持设置为“1小时后”、“2小时后”、“4小时后”、“6小时后”、“自定义”。如果选择“自定义”的模式，可在右侧输入框中输入1~24范围内的任意整数。
3. 启动部署上线后，可以在“部署上线”界面查看模型部署上线的状态。
部署上线将耗费较多时间，请您耐心等待。当“部署上线”页签版本管理区域的状态由“部署中”变更为“运行中”，部署完成。

📖 说明

在自动学习界面中，仅支持将训练后的模型部署为在线服务，如果需要部署为“批量服务”或“边缘服务”，请参见[自动学习生成的模型，存储在哪里？支持哪些其他操作？](#)。

服务测试

- 您可以在“部署上线”页面，选择对应的服务类型，例如自动学习物体检测项目默认将服务部署为在线服务，进入“在线服务”页面，单击目标服务“操作”列的“预测”，进行服务测试，测试方法和下方陈述操作步骤一致。具体操作请参见[测试服务](#)。
- 您也可以通过调用代码对服务进行测试，根据部署服务类型的不同，具体操作详情参见[访问在线服务](#)、[访问边缘服务](#)。
- 下面的测试，是您在自动学习物体检测项目页面将模型部署上线之后进行服务测试的操作步骤。
 - a. 模型部署完成后，您可添加图片进行测试。在“自动学习”页面，选择目标项目，进入“部署上线”界面，选择状态为“运行中”的服务版本，在“服务测试”区域单击“上传”，选择本地图片进行测试。

图 2-28 上传图片

服务测试

当服务状态为“✔ 运行中”时，才能进行预测。

选择需要预测的文件

上传

预测

- b. 单击“预测”进行测试，预测完成后，右侧“预测结果”区域输出结果。如模型准确率不满足预期，可在“数据标注”页签中添加图片并进行标注，重新进行模型训练及部署上线。预测结果中的参数说明请参见表2-10。如果您对模型预测结果满意，可根据界面提示调用接口访问在线服务，操作指导请参见“访问在线服务”。

目前只支持jpg、jpeg、bmp、png格式的图片。

图 2-29 预测结果

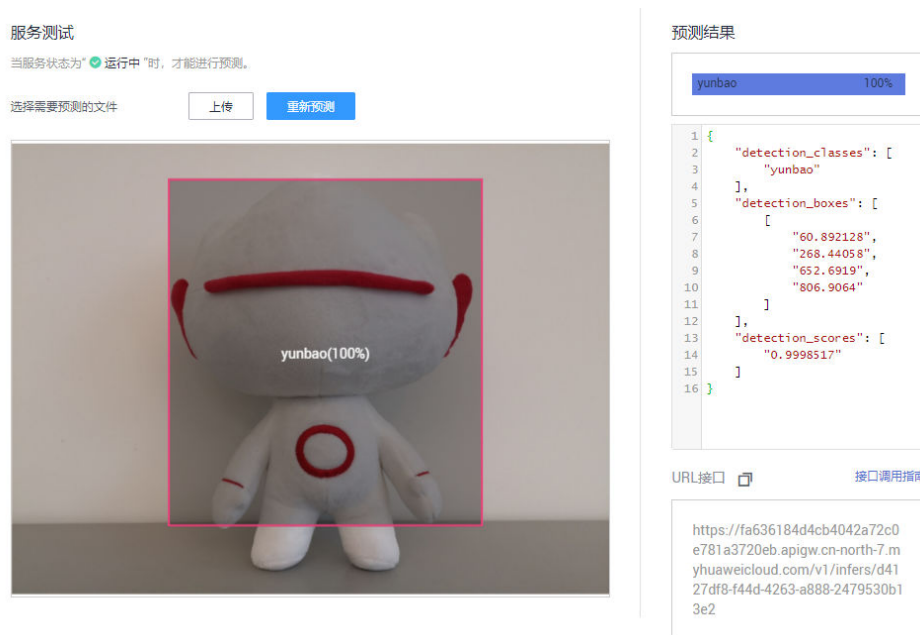
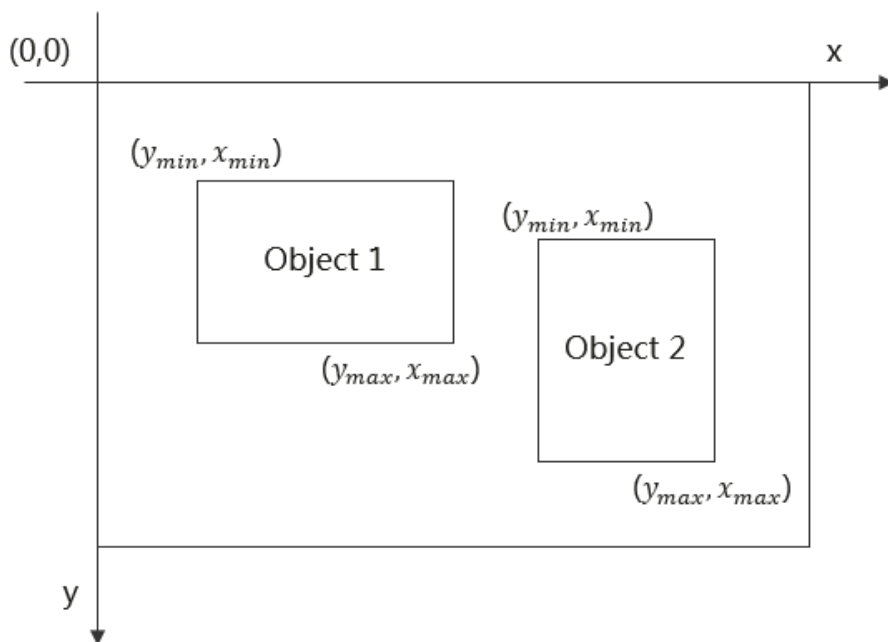


表 2-10 预测结果中的参数说明

参数	说明
detection_classes	每个检测框的标签。
detection_boxes	每个检测框的四点坐标 (y_min,x_min,y_max,x_max)，如图2-30所示。
detection_scores	每个检测框的置信度。

图 2-30 检测框的四点坐标示意图



说明

由于“运行中”的在线服务将持续耗费资源，如果不需再使用此在线服务，建议在版本管理区域，单击“停止”，即可停止在线服务的部署，避免产生不必要的费用。如果需要继续使用此服务，可单击“启动”恢复。

如果您启用了自动停止功能，服务将在指定时间后自动停止，不再产生费用。

2.4 预测分析

2.4.1 准备数据

使用ModelArts自动学习构建预测分析模型时，您需要将数据上传至对象存储服务（OBS）中。OBS桶需要与ModelArts在同一区域。

数据上传至 OBS

在本文档中，采用管理控制台上传数据至OBS。

执行如下操作，将数据导入到您的数据集中，以便用于模型训练和构建。

1. 登录OBS管理控制台，在ModelArts同一区域内**创建桶**。如果已存在可用的桶，需确保OBS桶与ModelArts在同一区域。
2. 参考**上传文件**，将本地数据上传至OBS桶中。如果您的数据较多，推荐OBS Browser+上传数据或上传文件夹。上传的数据需满足此类型自动学习项目的数据集要求。

说明

在上传数据时，请选择非加密桶进行上传，否则会由于加密桶无法解密导致后期的训练失败。

数据集要求

- 文件规范：名称由以字母数字及中划线下划线组成，以'.csv'结尾，且文件不能直接放在OBS桶的根目录下，应该存放在OBS桶的文件夹内。如：“/obs-xxx/data/input.csv”。
- 文件内容：文件保存为“csv”文件格式，文件内容以换行符（即字符“\n”，或称为LF）分隔各行，行内容以英文逗号（即字符“，”）分隔各列。文件内容不能包含中文字符，列内容不应包含英文逗号、换行符等特殊字符，不支持引号语法，建议尽量以字母及数字字符组成。
- 训练数据：
 - 训练数据列数一致，总数据量不少于100条不同数据（有一个特征取值不同，即视为不同数据）。
 - 训练数据列内容不能有时间戳格式（如：yy-mm-dd、yyyy-mm-dd等）的数据。
 - 如果某一列的取值只有一种，会被视为无效列。请确保标签列的取值至少有两个且无数据缺失。

说明

标签列指的是在训练任务中被指定为训练目标的列，即最终通过该数据集训练得到模型时的输出（预测项）。

- 除标签列外数据集中至少还应包含两个有效特征列（列的取值至少有两个且数据缺失比例低于10%）。
- 训练数据的csv文件不能包含表头，否则会导致训练失败。
- 当前由于特征筛选算法限制，标签列建议放在数据集最后一列，否则可能导致训练失败。

OBS 上传文件规范

预测分析项目的OBS数据路径需符合以下规则：

- 输入数据的OBS路径应指向数据文件，且文件不能直接放在OBS桶的根目录下，应该存放在OBS桶的文件夹内。如：“/obs-xxx/data/input.csv”。
- 输入数据的格式必须为csv格式，数据文件不包括表头，有效数据行数必须大于100行。列数必须小于200列，数据总大小不能超过100MB。

预测分析文件内容示例

以鸢尾花数据集为例：根据鸢尾花的花萼与花瓣长宽，预测鸢尾花的品种。

表 2-11 数据源的具体字段及意义

字段名	含义	类型	描述
attr_1	萼长	Double	花萼的长度
attr_2	萼宽	Double	花萼的宽度
attr_3	瓣长	Double	花瓣的长度
attr_4	瓣宽	Double	花瓣的宽度

字段名	含义	类型	描述
attr_5	品种	String	鸢尾花的品种

表 2-12 样本数据

attr_1	attr_2	attr_3	attr_4	attr_5
5.1	3.5	1.4	0.2	Iris-setosa
7	3.2	4.7	1.4	Iris-versicolor
6.3	3.3	6	2.5	Iris-virginica

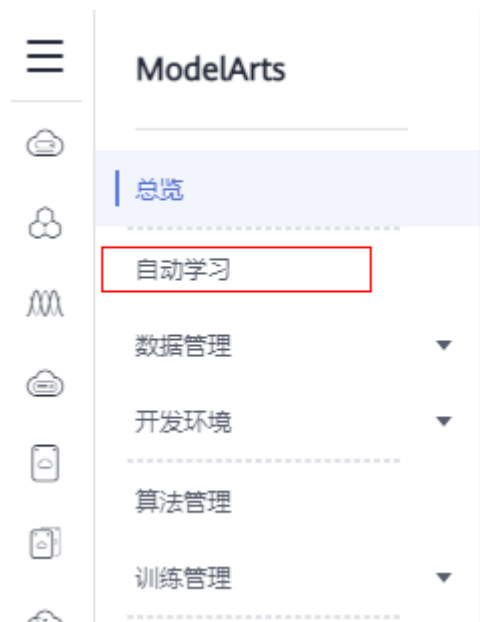
2.4.2 创建项目

ModelArts自动学习，包括图像分类项目、物体检测项目、预测分析项目、声音分类和文本分类项目。您可以根据业务需求选择创建合适的项目。您需要执行如下操作来创建自动学习项目。

创建项目

1. 登录ModelArts管理控制台，在左侧导航栏单击“自动学习”，单击“返回旧版”进入旧版“自动学习”页面。

图 2-31 进入自动学习



2. 在您需要的自动学习项目列表中，单击“创建项目”，进入创建自动学习项目界面。

图 2-32 自动学习列表



3. 在创建自动学习项目页面，计费模式默认“按需计费”，填写“名称”并选择“训练数据”的存储路径，训练数据路径选择已创建的OBS桶及文件夹，需指定至数据文件。

表 2-13 参数说明

参数	说明
“名称”	项目的名称。 <ul style="list-style-type: none"> 名称只能包含数字、字母、下划线和中划线，长度不能超过20位且不能为空。 名称请以大写或小写字母开头。
“训练数据”	OBS数据路径及数据文件，选择的OBS数据路径需满足规范。详情参见本章节 OBS上传文件规范 相关内容。 <ul style="list-style-type: none"> 除“准备数据>OBS上传文件规范”说明的文件及文件夹外，训练数据路径下不允许存任何其他文件或文件夹，否则会报错。 训练数据路径下的文件请不要随意修改。 <p>说明 每个训练数据OBS桶路径只能创建一个自动学习项目。如果需要对同一个数据集创建多个项目，则需要复制原OBS桶路径下数据到另一个OBS桶路径，然后再创建自动学习项目。</p>
“描述”	对项目的简要描述。

4. 单击“创建项目”，待页面提示“创建项目成功”完成创建，创建成功后页面自动跳转到“数据标注”。您也可以在“自动学习”页面查看“训练状态”为“未开始”的项目即为创建成功的项，单击项目名称进入“数据标注”。

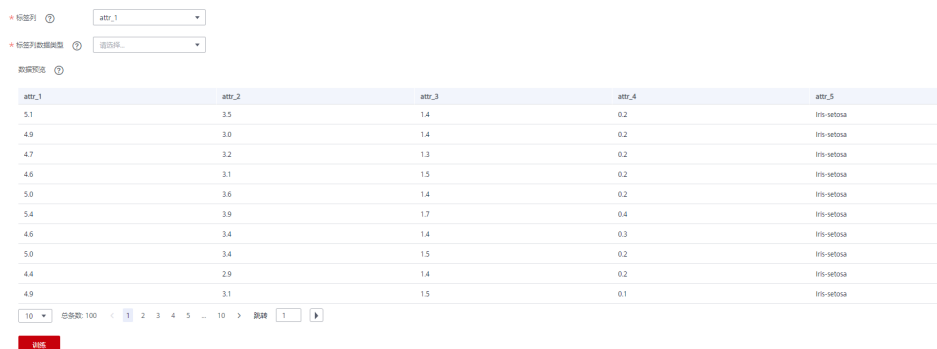
2.4.3 选择标签列

创建预测分析项目后，需要选择数据标签列及标签列数据类型。在预测分析“数据标注”页面，可预览数据并完成标签列以及标签列数据类型选择，当前由于特征筛选算法限制，标签列必须是数据集的最后一列。模型训练将会使用全部数据训练预测模型，该模型以其他列的数据为输入，以标签列的预测值为输出。

操作步骤

1. 选择“标签列”。在“数据标注”页面中，预览数据并选择训练目标。在“标签列”下拉框中选择需要设置为标签列的名称。
 标签列是预测模型的输出。此处训练目标是鸢尾花的品种（即“attr_5”），该列目标结果是“离散值”。训练目标选择完成后，单击“训练”。

图 2-33 预测分析数据标注界面



- 选择“标签列数据类型”。在“数据标注”页面中，“标签列数据类型”下选择合适的数据类型。
 - 若标签列为枚举型数据，数据类型应选择“离散值”，预测分析将训练分类模型。
 - 若标签列为数值型连续数据，数据类型应选择“连续数值”，预测分析将训练回归模型。

说明

- 分类问题（离散值）在模型训练完成后，评估结果会展现召回率（Recall）、精确率（Precision）、准确率（Accuracy）及F1值（F1 Score）。
- 连续值在模型训练完成后，评估结果会展现平均绝对误差（Mean Absolute Error），均方误差（Mean Squared Error），均方根误差（Root Mean Squared Error）。

2.4.4 模型训练

完成预测分析数据标注后，可进行模型的训练，得到预测分析的模型。部署上线步骤将使用预测模型发布在线预测服务。

操作步骤

- 在自动学习页面，单击创建成功的项目名称，进入“数据标注”页面，完成“标签列”和“标签列数据类型”的选择。
- 在“数据标注”页面下方，单击“训练”，在弹出的“训练设置”对话框中，选择训练使用的“计算规格”，单击“下一步”进入配置页，确认规格后单击“提交”开始模型训练。

预测分析类型的自动学习，只支持使用“自动学习专用实例CPU（8U）”训练模型。

训练时间相对较长，建议您耐心等待。如果关闭或退出此页面，系统仍然在执行训练操作。

图 2-34 训练设置

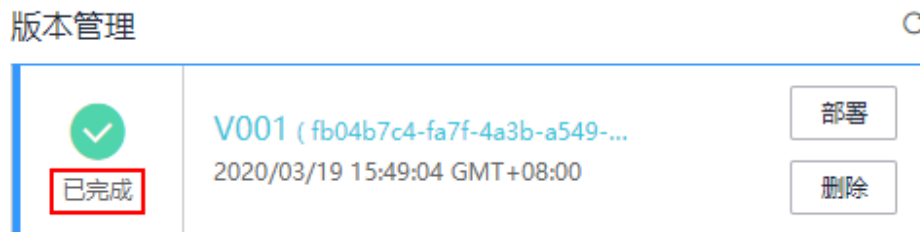
训练设置

计算规格

自动学习专用实例CPU（8U）

3. 在“模型训练”页签中，待训练状态由“运行中”变为“已完成”，即完成模型的自动训练。

图 2-35 运行成功



4. 训练完成后，您可以在界面中查看训练详情，如“标签列”和“标签列数据类型”、“准确率”、“评估结果”等。
该示例为二分类的离散型数值，评估效果参数说明请参见表2-14。
不同类型标签列数据产生的评估结果说明请参见[评估结果说明](#)。

说明

同一个自动学习项目可以训练多次，每次训练生成一个版本。如第一次训练版本号为“V001 (xxx)”，下一个版本为“V002 (xxx)”。基于训练版本可以对训练模型进行管理。当训练的模型达到目标后，再执行部署上线的操作。

评估结果说明

根据训练数据类的不同评估结果会包含不同的指标。

- 离散值评估结果
包含评估指标为召回率 (Recall)、精确率 (Precision)、准确率 (Accuracy) 与 F1值 (F1 Score)。下表为具体说明：

表 2-14 离散值评估结果包含指标说明

参数	说明
召回率	被用户标注为某个分类的所有样本中，模型正确预测为该分类的样本比率，反映模型对正样本的识别能力。
精确率	被模型预测为某个分类的所有样本中，模型正确预测的样本比率，反映模型对负样本的区分能力。
准确率	所有样本中，模型正确预测的样本比率，反映模型对样本整体的识别能力。
F1值	F1值是模型精确率和召回率的加权调和平均，用于评价模型的好坏，当F1较高时说明模型效果较好。

- 连续数值评估结果
包含评估指标为平均绝对误差 (Mean Absolute Error)、均方误差 (Mean Squared Error) 与均方根误差 (Root Mean Squared Error)。三个误差值能够表征真实值和预测值之间的差距。在多次建模的过程中，每一次建模结果都会产

生一组误差值，评判一个模型好坏的方法就是看这三个误差值是否变小或者变大，误差值越小表示模型越好。

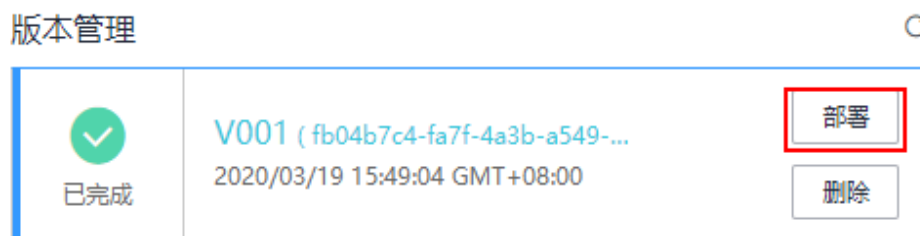
2.4.5 部署上线

部署上线

部署上线操作即将模型部署为在线服务，并且提供在线的测试UI与监控能力。完成模型训练后，可选择准确率理想且训练状态为“运行成功”的版本部署上线。具体操作步骤如下。

1. 在“模型训练”页签中，待训练状态变为“运行成功”，单击版本管理区域中的“部署”。

图 2-36 部署操作

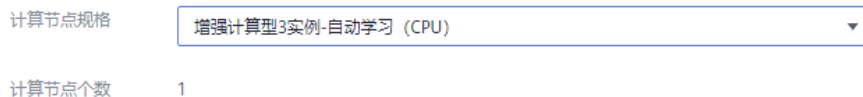


2. 在弹出的“部署”对话框中，选择部署上线使用的“资源规格”，然后单击“确定”开始将模型部署上线为在线服务。
 - “计算节点规格”：选择计算资源，以控制台实际提供的规格为准。
 - “计算节点个数”：默认为1，且不能修改。
 - “是否自动停止”：启用该参数并设置时间后，服务将在指定时间后自动停止。如果不启用此参数，在线服务将一直运行，同时一直收费，自动停止功能可以帮您避免产生不必要的费用。默认开启自动停止功能，且默认值为“1小时后”。

目前支持设置为“1小时后”、“2小时后”、“4小时后”、“6小时后”、“自定义”。如果选择“自定义”的模式，可在右侧输入框中输入1~24范围内的任意整数。

图 2-37 部署上线

部署



3. 启动部署上线后，可以在“部署上线”界面查看模型部署上线的状态。部署上线将耗费较多时间，请您耐心等待。当“部署上线”页签版本管理区域的状态由“部署中”变更为“运行中”，部署完成。

说明

在自动学习界面中，仅支持将训练后的模型部署为在线服务，如果需要部署为“批量服务”或“边缘服务”，请参见[自动学习生成的模型，存储在哪里？支持哪些其他操作？](#)。

服务测试

- 您可以在“部署上线”页面，选择对应的服务类型，例如自动学习预测分析项目默认将服务部署为在线服务，进入“在线服务”页面，单击目标服务“操作”列的“预测”，进行服务测试，测试方法和下方陈述操作步骤一致。具体操作请参见[测试服务](#)。
 - 您也可以通过调用代码对服务进行测试，根据部署服务类型的不同，具体操作详情参见[访问在线服务](#)、[访问边缘服务](#)。
 - 下面的测试，是您在自动学习预测分析项目页面将模型部署上线之后进行服务测试的操作步骤。
 - a. 模型部署完成后，您可输入代码进行测试。在“自动学习”页面，选择目标项目，进入“部署上线”界面，选择状态为“运行中”的服务版本，在“服务测试”的“代码”区域，输入调试代码。
 - b. 单击“预测”进行测试，预测完成后，右侧“返回结果”区域输出测试结果。如模型准确率不满足预期，可在“数据标注”页签，重新进行模型训练及部署上线。如果您对模型预测结果满意，可根据界面提示调用接口访问在线服务，操作指导请参见“[访问在线服务](#)”。
- 输入代码：attr_1~attr_4为输入的预测数据，其中在“数据标注”中，选择的标注列为attr_5，即attr_5为需要预测的目标列，输入数据中不再需要填写。

```
{
  "data": {
    "req_data": [{
      "attr_1": 5.1,
      "attr_2": 3.5,
      "attr_3": 1.4,
      "attr_4": 0.2
    }]
  }
}
```

- 返回结果如[图2-38](#)所示：predict为目标列attr_5的预测结果。

图 2-38 预测结果



说明

由于“运行中”的在线服务将持续耗费资源，如果不再使用此在线服务，建议在版本管理区域，单击“停止”，即可停止在线服务的部署，避免产生不必要的费用。如果需要继续使用此服务，可单击“启动”恢复。

2.5 声音分类

2.5.1 准备数据

使用ModelArts自动学习构建模型时，您需要将数据上传至对象存储服务（OBS）中。OBS桶需要与ModelArts在同一区域。

数据上传至 OBS

在本文档中，采用管理控制台上传数据至OBS。

执行如下操作，将数据导入到您的数据集中，以便用于模型训练和构建。

1. 登录OBS管理控制台，在ModelArts同一区域内**创建桶**。如果已存在可用的桶，需确保OBS桶与ModelArts在同一区域。
2. 参考**上传文件**，将本地数据上传至OBS桶中。如果您的数据较多，推荐OBS Browser+上传数据或上传文件夹。上传的数据需满足此类型自动学习项目的数据集要求。

说明

在上传数据时，请选择非加密桶进行上传，否则会由于加密桶无法解密导致后期的训练失败。

声音分类的数据要求

- 音频只支持16bit的WAV格式。支持WAV的所有子格式。
- 单条音频时长应大于1s，大小不能超过4MB。
- 适当增加训练数据，会提升模型的精度。声音分类建议每类音频至少50条，每类音频总时长至少5分钟。
- 建议训练数据和真实识别场景的声音保持一致并且每类的音频尽量覆盖真实环境的所有场景。
- 训练集的数据质量对于模型的精度有很大影响，建议训练集音频的采样率和采样精度保持一致。
- 标注质量对于最终的模型精度有极大的影响，标注过程中尽量不要出现误标情况。
- 音频标注涉及到的标注标签和声音内容只支持中文和英文，不支持小语种。

OBS 上传文件规范

- 如不需要提前上传训练数据，请创建一个空文件夹用于存放工程后期生成的文件。如：“/bucketName/data-cat”。
- 如需要提前上传待标注的音频，请创建一个空文件夹，然后将音频文件保存在该文件夹下，音频的目录结构如：“/bucketName/data-cat/cat.wav”。

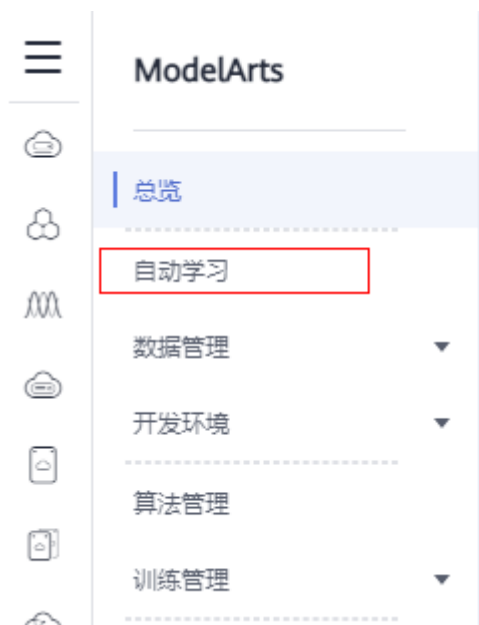
2.5.2 创建项目

ModelArts自动学习，包括图像分类项目、物体检测项目、预测分析项目、声音分类和文本分类项目。您可以根据业务需求选择创建合适的项目。您需要执行如下操作来创建自动学习项目。

创建项目

1. 登录ModelArts管理控制台，在左侧导航栏单击“自动学习”，单击“返回旧版”进入旧版“自动学习”页面。

图 2-39 进入自动学习



2. 在您需要的自动学习项目列表中，单击“创建项目”，进入创建自动学习项目界面。

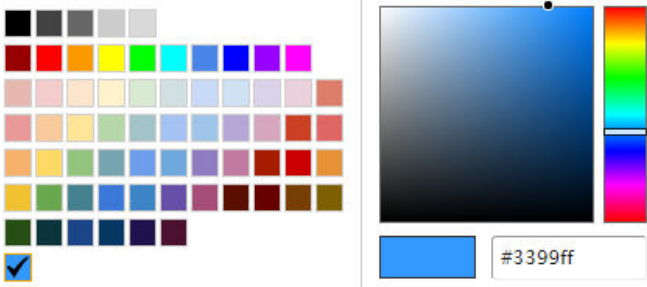
图 2-40 自动学习列表



3. 在创建自动学习项目页面，计费模式默认“按需计费”，参考表2-15填写相应参数。

表 2-15 参数说明

参数	说明
“名称”	项目的名称。 <ul style="list-style-type: none"> 名称只能包含数字、字母、下划线和中划线，长度不能超过32位且不能为空。 名称请以大写或小写字母开头。
“描述”	对项目的简要描述。

参数	说明
“数据集来源”	<p>可选择“新建数据集”或“已有数据集”。</p> <ul style="list-style-type: none"> “新建数据集”：需填写创建数据集所需的参数，包含“数据集名称”、“数据集输入位置”、“数据集输出位置”和“添加标签集”信息。 “已有数据集”：您可以从ModelArts数据管理中选择一个同类型的数据集，用于创建自动学习项目。“数据集名称”的下拉框中，仅展示同类型的数据集供选择。
“数据集名称”	<p>当选择“新建数据集”时，此参数右侧为文本框，需根据规则填写新创建的数据集名称；当选择“已有数据集”时，此参数右侧为下拉框，展示当前账号下可用的同类型数据集列表。</p>
“数据集输入位置”	<p>选择数据集输入位置的OBS路径。数据集输入的规范请参见准备数据章节。</p> <ul style="list-style-type: none"> 除“准备数据>OBS上传文件规范”说明的文件及文件夹外，训练数据路径下不允许存任何其他文件或文件夹，否则会报错。 训练数据路径下的文件请不要随意修改。
“数据集输出位置”	<p>选择数据集输出位置的OBS路径。</p> <p>说明 “数据集输出位置”不能与“数据集输入位置”为同一路径，且不能是“数据集输入位置”的子目录。“数据集输出位置”建议选择一个空目录。</p>
“添加标签集”	<ul style="list-style-type: none"> 设置标签名称：在标签名称文本框中，输入标签名称。标签名称只能是中文、字母、数字、下划线或中划线组成的合法字符串。长度为1~32字符。 添加标签：单击“添加标签”可增加多个标签。 设置标签颜色：“物体检测”和“文本分类”的数据集需设置标签颜色，“图像分类”和“声音分类”类型不需要。在每个标签右侧的标签颜色区域下，可在色板中选择颜色，或者直接输入十六进制颜色码进行设置。 

- 单击“创建项目”，待页面提示“创建项目成功”完成创建，创建成功后页面自动跳转到“数据标注”。您也可以直接在“自动学习”页面查看创建成功的项目，单击项目名称进入“数据标注”。

2.5.3 数据标注

音频标注

1. 选择未标注音频。在“数据标注”页面单击“未标注”页签，此页面展示所有未标注的音频数据。依次单击选中待标注的音频，或勾选“选择当前页”选中该页面所有音频，在页面右侧进行标注。

图 2-41 音频标注



2. 添加标注。先对音频进行播放识别，然后选中音频文件，在右侧“标签”区域，输入“标签名”或从下拉列表中选择已添加的标签，同时可在下拉菜单中选择标签“快捷键”。单击“确定”，完成选中音频的标注操作。

图 2-42 声音分类音频标注



3. 当目录中所有音频都完成标注后，您可以在“已标注”页签下查看已完成标注的音频，或者通过右侧的“全部标签”列表，了解当前已完成的标签名称和标签数量。

同步或添加音频

在“自动学习”页面，单击项目名称，进入自动学习“数据标注”页面。声音分类项目创建时，音频来源有两种，通过本地添加或同步OBS中的数据。

- **添加音频：**您可以将本地音频快速添加到ModelArts，同时自动上传至创建项目时所选择的OBS路径中。单击“添加音频”，在弹出的对话框中单击“添加音频”并添加。

说明

仅支持16bit WAV格式音频文件，单个音频文件不能超过4MB，且单上传的音频文件总大小不能超过8MB。

- **数据源同步：**为了快速获取用户OBS桶中最新音频，单击“数据源同步”，快速将通过OBS上传的音频数据添加到ModelArts。
- **删除音频：**您可以依次单击选中音频，或勾选“选择当前页”选中该页面所有音频进行删除操作。

📖 说明

所有的删除操作均不可恢复，请谨慎操作。

修改标注

当数据完成标注后，您还可以进入“已标注”页签，对已标注的数据进行修改。

• 基于音频修改

在数据集详情页，单击“已标注”页签，然后在音频列表中选中待修改的音频（选择一个或多个）。在右侧标签信息区域中对标签进行修改。

- 修改标签：在“选中文件标签”区域中，单击操作列的编辑图标，然后在文本框中输入正确的标签名，然后单击确定图标完成修改。
- 删除标签：在“选中文件标签”区域中，单击操作列的删除图标，在弹出的对话框中单击“确定”删除该标签。

• 基于标签修改

在数据标注页面，单击“标签管理”页签，在标签管理页，显示全部标签的信息。

图 2-43 全部标签的信息



2.5.4 模型训练

完成音频标注后，可进行模型的训练。模型训练的目的在于得到满足需求的声音分类模型。由于用于训练的音频，至少有2种以上的分类，每种分类的音频数不少于5个。因此在发布训练之前，请确保已标注的音频符合要求，否则下方“开始训练”按钮会处于灰色状态。

操作步骤

在开始训练之前，需要设置训练参数，然后再开始模型的自动训练。

1. 在自动学习页面，单击创建成功的项目名称，进入“数据标注”页面，完成数据标注。

图 2-44 完成数据标注



2. 在“数据标注”页面，单击右上角“开始训练”，然后在弹出的“训练设置”对话框中，参考表2-16填写相关参数，然后单击“确定”，开始进行模型训练。

表 2-16 训练设置参数说明

参数	说明	默认值
数据集版本名称	此版本即数据管理中发布数据集时设置的版本。自动学习项目中，启动训练作业时，会基于前面的数据标注，将数据集发布为一个版本。系统将自动给出一个版本号，您也可以根据实际情况进行填写。	系统随机给出
最大训练时长（分钟）	设置最大训练时长，在该时长内若训练还未完成，则强制退出。为防止训练中退出，建议使用较大值。输入取值范围为6~6000。	60
计算规格	选择训练使用的资源规格。	以控制台实际提供为准。

- 训练参数设置完成后，单击“下一步”进入配置页，确认规格后单击“提交”进行模型的自动训练，训练时间相对较长，建议您耐心等待。如果关闭或退出此页面，系统仍然在执行训练操作。
- 在“模型训练”页签中，待训练状态由“运行中”变为“已完成”，即完成模型的自动训练。
- 训练完成后，您可以在界面中查看训练详情，如“准确率”、“评估结果”、“训练参数”、“分类统计表”等。

图 2-45 训练详情



表 2-17 评估结果参数说明

参数	说明
召回率	被用户标注为某个分类的所有样本中，模型正确预测为该分类的样本比率，反映模型对正样本的识别能力。
精确率	被模型预测为某个分类的所有样本中，模型正确预测的样本比率，反映模型对负样本的区分能力。

参数	说明
准确率	所有样本中，模型正确预测的样本比率，反映模型对样本整体的识别能力。
F1值	F1值是模型精确率和召回率的加权调和平均，用于评价模型的好坏，当F1较高时说明模型效果较好。

📖 说明

同一个自动学习项目可以训练多次，每次训练生成一个版本。如第一次训练版本号为“V001 (xxx)”，下一个版本为“V002 (xxx)”。基于训练版本可以对训练模型进行管理。当训练的模型达到目标后，再执行部署上线的操作。

2.5.5 部署上线

部署上线

部署上线操作即将模型部署为在线服务，并且提供在线的测试UI与监控能力。完成模型训练后，可选择准确率理想且训练状态为“运行成功”的版本部署上线。具体操作步骤如下。

1. 在“模型训练”页签中，待训练状态变为“运行成功”，单击版本管理区域中的“部署”，开始将模型部署上线为在线服务。

图 2-46 部署操作



2. 在弹出的“部署”对话框中，选择资源规格，同时设置自动停止功能，然后单击确定，启动部署。
 - “计算节点规格”：以控制台实际提供为准。
 - “计算节点个数”：默认为1，且不能修改。
 - “是否自动停止”：启用该参数并设置时间后，服务将在指定时间后自动停止。如果不启用此参数，在线服务将一直运行，同时一直收费，自动停止功能可以帮您避免产生不必要的费用。默认开启自动停止功能，且默认值为“1小时后”。

目前支持设置为“1小时后”、“2小时后”、“4小时后”、“6小时后”、“自定义”。如果选择“自定义”的模式，可在右侧输入框中输入1~24范围内的任意整数。

3. 启动部署上线后，可以在“部署上线”界面查看模型部署上线的状态。部署上线将耗费较多时间，请您耐心等待。当“部署上线”页签版本管理区域的状态由“部署中”变更为“运行中”，部署完成。

 说明

在自动学习界面中，仅支持将训练后的模型部署为在线服务，如果需要部署为“批量服务”或“边缘服务”，请参见[自动学习生成的模型，存储在哪里？支持哪些其他操作？](#)。

服务测试

- 您可以在“部署上线”页面，选择对应的服务类型，例如自动学习声音分类项目默认将服务部署为在线服务，进入“在线服务”页面，单击目标服务“操作”列的“预测”，进行服务测试，测试方法和下方陈述操作步骤一致。具体操作请参见[测试服务](#)。
- 您也可以通过调用代码对服务进行测试，根据部署服务类型的不同，具体操作详情参见[访问在线服务](#)、[访问边缘服务](#)。
- 下面的测试，是您在自动学习声音分类项目页面将模型部署上线之后进行服务测试的操作步骤。
 - 模型部署完成后，您可添加音频文件进行测试。在“自动学习”页面，选择目标项目，进入“部署上线”界面，选择状态为“运行中”的服务版本，在“服务测试”区域单击“上传”，选择本地音频进行测试。
 - 单击“预测”进行测试，预测完成后，右侧“预测结果”区域输出测试结果。如模型准确率不满足预期，可在“数据标注”页签中添加音频并进行标注，重新进行模型训练及部署上线。预测结果中的参数说明请参见[表2-18](#)。如果您对模型预测结果满意，可根据界面提示调用接口访问在线服务，操作指导请参见“[访问在线服务](#)”。

表 2-18 预测结果中的参数说明

参数	说明
predicted_label	该段音频的预测类别。
score	预测为此类别的置信度。

 说明

由于“运行中”的在线服务将持续耗费资源，如果不需再使用此在线服务，建议在版本管理区域，单击“停止”，即可停止在线服务的部署，避免产生不必要的费用。如果需要继续使用此服务，可单击“启动”恢复。

如果您启用了自动停止功能，服务将在指定时间后自动停止，不再产生费用。

2.6 文本分类

2.6.1 准备数据

使用ModelArts自动学习构建模型时，您需要将数据上传至对象存储服务（OBS）中。OBS桶需要与ModelArts在同一区域。

数据上传至 OBS

在本文档中，采用管理控制台上传数据至OBS。

执行如下操作，将数据导入到您的数据集中，以便用于模型训练和构建。

1. 登录OBS管理控制台，在ModelArts同一区域内**创建桶**。如果已存在可用的桶，需确保OBS桶与ModelArts在同一区域。
2. 参考**上传文件**，将本地数据上传至OBS桶中。如果您的数据较多，推荐OBS Browser+上传数据或上传文件夹。上传的数据需满足此类型自动学习项目的数据集要求。

📖 说明

在上传数据时，请选择非加密桶进行上传，否则会由于加密桶无法解密导致后期的训练失败。

数据集要求

- 文件格式要求为txt或者csv，文件大小不能超过8MB。
- 以换行符作为分隔符，每行数据代表一个标注对象。
- 文本分类目前只支持中文。

OBS 上传文件规范

- 如不需要提前上传训练数据，请创建一个空文件夹用于存放工程后期生成的文件。
- 如需要提前上传待标注的文件，请创建一个空文件夹，然后将文本文件保存在该文件夹下，文本文件的目录结构如：“/bucketName/data/text.csv”。
- 标签名是由中文、大小写字母、数字、中划线或下划线组成，且不超过32位的字符串。
- 如您将已标注好的文本文件上传至OBS桶，请按照如下规范上传。
 - 要求将标注对象和标注文件存储在同一目录，并且一一对应，如标注对象文件名为“COMMENTS_114745.txt”，那么标注文件名为“COMMENTS_114745_result.txt”。

数据文件存储示例：

```
|<dataset-import-path>
|
|   COMMENTS_114732.txt
|   COMMENTS_114732_result.txt
|   COMMENTS_114745.txt
|   COMMENTS_114745_result.txt
|   COMMENTS_114945.txt
|   COMMENTS_114945_result.txt
```

- 文本分类的标注对象和标注文件均为文本文件，并且以行数进行对应。如标注文件中的第一行表示的是标注对象文件中的第一行的标注。

例如，标注对象“COMMENTS_114745.txt”的内容如下所示。

```
手感很好，反应速度很快，不知道以后怎样
三个月前买了一个用的非常好果断把旧手机替换下来尤其在待机方面
没充一会电源怎么也会发热呢音量键不好用回弹不好
算是给自己的父亲节礼物吧物流很快下单不到24小时就到货了耳机更赞有些低音炮的感觉入耳很紧不会掉棒棒哒
```

标注文件“COMMENTS_114745_result.txt”的内容。

```
positive
positive
negative
positive
```

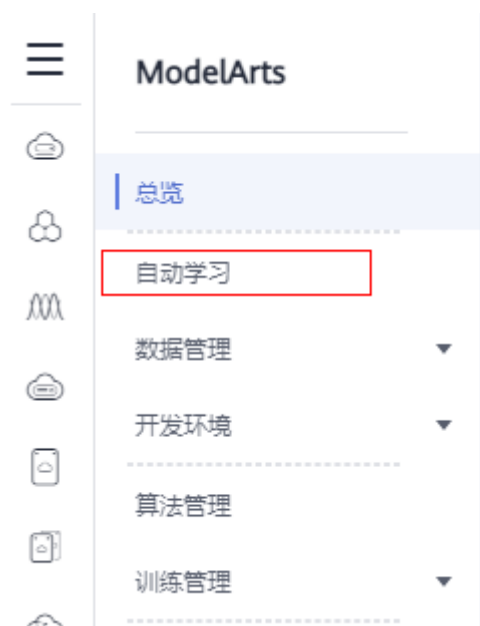
2.6.2 创建项目

ModelArts自动学习，包括图像分类项目、物体检测项目、预测分析项目、声音分类和文本分类项目。您可以根据业务需求选择创建合适的项目。您需要执行如下操作来创建自动学习项目。

创建项目

1. 登录ModelArts管理控制台，在左侧导航栏单击“自动学习”，单击“返回旧版”进入旧版“自动学习”页面。

图 2-47 进入自动学习



2. 在您需要的自动学习项目列表中，单击“创建项目”，进入创建自动学习项目界面。

图 2-48 自动学习列表



3. 在创建自动学习项目页面，计费模式默认“按需计费”，参考表2-19填写相应参数。

表 2-19 参数说明

参数	说明
“名称”	项目的名称。 <ul style="list-style-type: none"> 名称只能包含数字、字母、下划线和中划线，长度不能超过32位且不能为空。 名称请以大写或小写字母开头。
“描述”	对项目的简要描述。
“数据集来源”	可选择“新建数据集”或“已有数据集”。 <ul style="list-style-type: none"> “新建数据集”：需填写创建数据集所需的参数，包含“数据集名称”、“数据集输入位置”、“数据集输出位置”和“添加标签集”信息。 “已有数据集”：您可以从ModelArts数据管理中选择一个同类型的数据集，用于创建自动学习项目。“数据集名称”的下拉框中，仅展示同类型的数据集供选择。
“数据集名称”	当选择“新建数据集”时，此参数右侧为文本框，需根据规则填写新创建的数据集名称；当选择“已有数据集”时，此参数右侧为下拉框，展示当前账号下可用的同类型数据集列表。
“数据集输入位置”	选择数据集输入位置的OBS路径。数据集输入的规范请参见 准备数据 章节。 <ul style="list-style-type: none"> 除“准备数据>OBS上传文件规范”说明的文件及文件夹外，训练数据路径下不允许存任何其他文件或文件夹，否则会报错。 训练数据路径下的文件请不要随意修改。
“数据集输出位置”	选择数据集输出位置的OBS路径。 说明 “数据集输出位置”不能与“数据集输入位置”为同一路径，且不能是“数据集输入位置”的子目录。“数据集输出位置”建议选择一个空目录。

参数	说明
“添加标签集”	<ul style="list-style-type: none"> ● 设置标签名称: 在标签名称文本框中, 输入标签名称。标签名称只能是中文、字母、数字、下划线或中划线组成的合法字符串。长度为1~32字符。 ● 添加标签: 单击“添加标签”可增加多个标签。 ● 设置标签颜色: “物体检测”和“文本分类”的数据集需设置标签颜色, “图像分类”和“声音分类”类型不需要。在每个标签右侧的标签颜色区域下, 可在色板中选择颜色, 或者直接输入十六进制颜色码进行设置。 

4. 单击“创建项目”, 待页面提示“创建项目成功”完成创建, 创建成功后页面自动跳转到“数据标注”。您也可以直接在“自动学习”页面查看创建成功的项目, 单击项目名称进入“数据标注”。

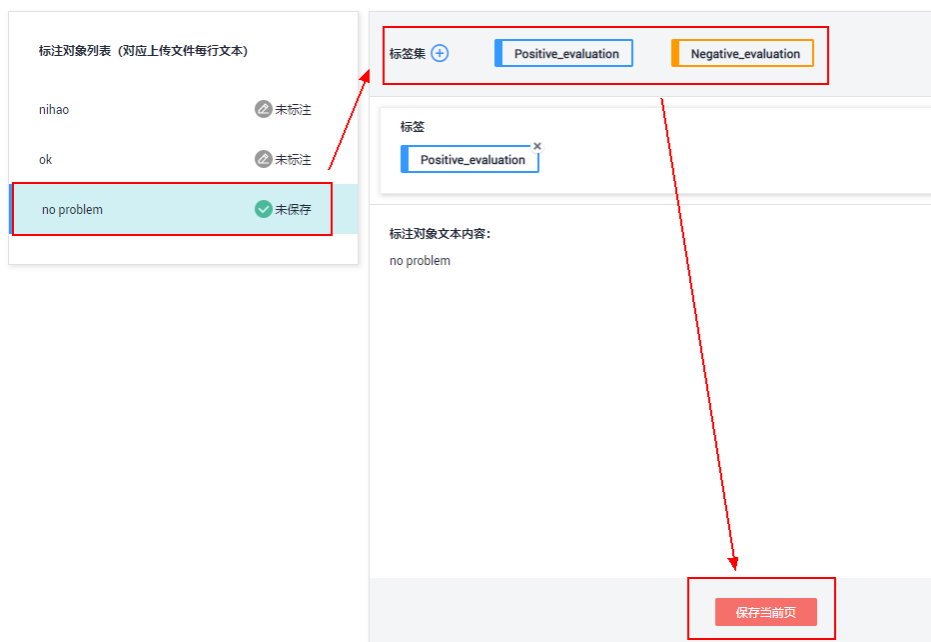
2.6.3 数据标注

文本分类项目创建完成后, 将自动跳转至“自动学习 > 数据标注”页面, 界面默认“已标注”页签。当选择的数据集已存在标注好的数据时, 将自动展示已标注数据。您也可以单击“未标注”页签, 切换至“未标注”页面, 展示数据集的输入目录中的未标注数据。

文本分类的数据标注

1. 在“标注对象列表”中选中待标注文本, 然后单击“标签集”区域中不同标签进行标注。
一个文本对象只能添加一个标签。
2. 确认文件标签后, 单击右下方“保存当前页”, 完成标注。
当“标注对象列表”内容较多时, 其区域下方将呈现翻页, 请务必在本页完成标注后, 单击“保存当前页”保存后再翻页。如果您未完成保存即翻页, 将导致前一页的标注信息丢失, 需重新标注。

图 2-49 数据标注-文本分类



添加或删除数据

自动学习项目中，数据来源为数据集中输入位置对应的OBS目录，当目录下的数据无法满足现有业务时，您可以在ModelArts自动学习页面中，添加或删除数据。

- **添加文件**

在“未标注”页签下，可单击页面左上角的“添加文件”，您可以在弹出对话框中，选择本地文件上传。

上传文件格式需满足文本分类的[数据集要求](#)。

- **删除文本对象**

在“已标注”页签下或“未标注”页签下，选中需要删除的文本对象，单击页面左上角的“删除”，在弹出的对话框中，确认删除信息后，单击“确定”。

在“已标注”页签下，您还可以勾选“选择当前页”，单击“删除”，即可删除当前页下所有的文本对象及其标注信息。

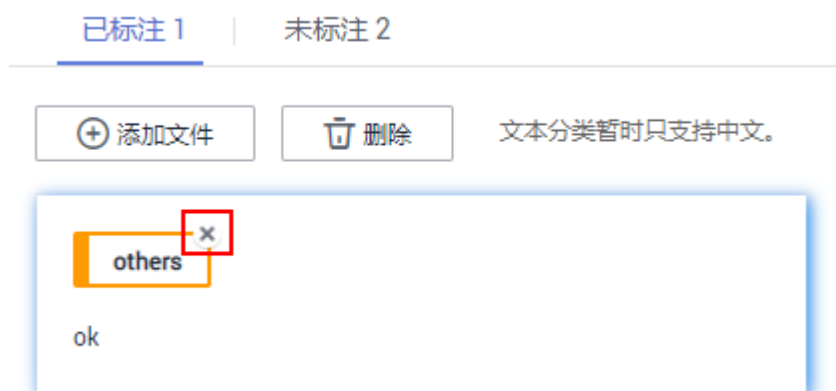
图 2-50 添加文件或删除文本对象



修改已标注的数据

针对“已标注”的文本数据，仅支持删除此文本对象的标签。在“已标注”页签下，单击标签右上角的叉号，即可删除此文本对象的标签。在弹出的对话框中，确认信息无误后单击“确定”。标签删除后，此文本对象将被呈现至“未标注”页签下。

图 2-51 删除已标注文本的标签



修改标签

针对文本分类的自动学习项目，项目创建成功后，您可以根据业务变化，修改用于标注的标签。支持添加、修改和删除标签。

- 添加标签
在“已标注”页签下，单击“全部标签”右侧的加号，在弹出“新增标签”对话框中，设置“标签名称”和“标签颜色”，然后单击“确定”完成标签添加。
- 修改标签
在“已标注”页签中“全部标签”的下方，选择需要修改的标签，单击操作列的编辑图标，在弹出“修改标签”对话框中，修改“标签名称”或“标签颜色”，然后单击“确定”完成标签修改。
- 删除标签
在“已标注”页签中“全部标签”的下方，选择需要删除的标签，单击操作列的删除图标，在弹出“删除”对话框中，选择“仅删除标签”或“删除标签及仅包含此标签的标注对象”，然后单击“确定”完成标签删除。

说明

所有的删除操作均不可恢复，请谨慎操作。

图 2-52 修改标签



2.6.4 模型训练

完成数据标注后，可进行模型的训练。模型训练的目的在于得到满足需求的文本分类模型。由于用于训练的文本，至少有2种以上的分类（即2种以上的标签），每种分类的文本数不少于20个。因此在发布训练之前，请确保已标注的文本符合要求，否则下方“开始训练”按钮会处于灰色状态。

操作步骤

1. 在“自动学习”页面，单击创建成功的项目名称，进入“数据标注”页面，完成数据标注。

图 2-53 完成数据标注



2. 在“数据标注”页面，单击右上角的“开始训练”，然后在弹出的“训练设置”对话框中配置相关参数。参数说明请参见表2-20。

表 2-20 训练设置参数说明

参数	说明	默认值
数据集版本名称	此版本即数据管理中发布数据集时设置的版本。自动学习项目中，启动训练作业时，会基于前面的数据标注，将数据集发布为一个版本。 系统将自动给出一个版本号，您可以根据实际情况进行填写。	系统随机给出
训练验证比例	训练验证比例表示将已标注样本随机分为训练集和验证集的比例，默认训练集比例为0.8，即全部分为训练集，manifest中的usage字段记录划分类别。取值范围为0~1。	0.8
最大训练时长（分钟）	即最大训练时长，在该时长内若训练还未完成，则强制退出。为防止训练中退出，建议使用较大值。输入范围为6~6000。适当延长训练时间，500条文本的训练集建议选择运行120分钟以上。	60
训练偏好	<ul style="list-style-type: none"> performance_first: 性能优先，训练时间较短，模型较小 balance: 平衡 accuracy_first: 精度优先，训练时间较长，模型较大 	balance
计算规格	选择训练使用的资源规格。	以控制台实际为准。

3. 训练参数设置完成后，单击“下一步”进入配置页，确认规格后单击“提交”进行模型的自动训练，训练时间相对较长，建议您耐心等待。如果关闭或退出此页面，系统仍然在执行训练操作。
4. 在“模型训练”页签中，待训练状态由“运行中”变为“已完成”，即完成模型的自动训练。
5. 训练完成后，您可以在界面中查看训练详情，如“准确率”、“评估结果”、“训练参数”、“分类统计表”等。评估结果参数说明请参见表2-21。

图 2-54 训练详情



表 2-21 评估结果参数说明

参数	说明
召回率	被用户标注为某个分类的所有样本中，模型正确预测为该分类的样本比率，反映模型对正样本的识别能力。
精确率	被模型预测为某个分类的所有样本中，模型正确预测的样本比率，反映模型对负样本的区分能力。
准确率	所有样本中，模型正确预测的样本比率，反映模型对样本整体的识别能力。
F1值	F1值是模型精确率和召回率的加权调和平均，用于评价模型的好坏，当F1较高时说明模型效果较好。

说明

同一个自动学习项目可以训练多次，每次训练生成一个版本。如第一次训练版本号为“V001 (xxx)”，下一个版本为“V002 (xxx)”。基于训练版本可以对训练模型进行管理。当训练的模型达到目标后，再执行部署上线的操作。

2.6.5 部署上线

部署上线

部署上线操作即将模型部署为在线服务，并且提供在线的测试UI与监控能力。完成模型训练后，可选择准确率理想且训练状态为“运行成功”的版本部署上线。具体操作步骤如下。

1. 在“模型训练”页签中，待训练状态变为“运行成功”，单击版本管理区域中的“部署”，开始将模型部署上线为在线服务。

图 2-55 部署操作



2. 在弹出的“部署”对话框中，选择资源规格，同时设置自动停止功能，然后单击确定，启动部署。
 - “计算节点规格”：以控制台实际为准。
 - “计算节点个数”：默认为1，且不能修改。
 - “是否自动停止”：启用该参数并设置时间后，服务将在指定时间后自动停止。如果不启用此参数，在线服务将一直运行，同时一直收费，自动停止功能可以帮您避免产生不必要的费用。默认开启自动停止功能，且默认值为“1小时后”。

目前支持设置为“1小时后”、“2小时后”、“4小时后”、“6小时后”、“自定义”。如果选择“自定义”的模式，可在右侧输入框中输入1~24范围内的任意整数。

3. 启动部署上线后，可以在“部署上线”界面查看模型部署上线的状态。
部署上线将耗费较多时间，请您耐心等待。当“部署上线”页签版本管理区域的状态由“部署中”变更为“运行中”，部署完成。

📖 说明

在自动学习界面中，仅支持将训练后的模型部署为在线服务，如果需要部署为“批量服务”或“边缘服务”，请参见[自动学习生成的模型，存储在哪里？支持哪些其他操作？](#)。

服务测试

- 您可以在“部署上线”页面，选择对应的服务类型，例如自动学习文本分类项目默认将服务部署为在线服务，进入“在线服务”页面，单击目标服务“操作”列的“预测”，进行服务测试，测试方法和下方陈述操作步骤一致。具体操作请参见[测试服务](#)。
- 您也可以通过调用代码对服务进行测试，根据部署服务类型的不同，具体操作详情参见[访问在线服务](#)、[访问边缘服务](#)。
- 下面的测试，是您在自动学习文本分类项目页面将模型部署上线之后进行服务测试的操作步骤。
 - a. 模型部署完成后，您可添加文本进行测试。在“自动学习”页面，选择目标项目，进入“部署上线”界面，选择状态为“运行中”的服务版本，在“服务测试”区域的文本框中，输入需测试的文本。
 - b. 单击“预测”进行测试，预测完成后，右侧“预测结果”区域输出测试结果。如模型准确率不满足预期，可在“数据标注”页签中添加数据并进行标注，重新进行模型训练及部署上线。预测结果中的参数说明请参见[表2-22](#)。如果您对模型预测结果满意，可根据界面提示调用接口访问在线服务，操作指导请参见[“访问在线服务”](#)。

图 2-56 预测



表 2-22 预测结果中的参数说明

参数	说明
predicted_label	该段文本的预测类别。
score	预测为此类别的置信度。

说明

由于“运行中”的在线服务将持续耗费资源，如果不需再使用此在线服务，建议在版本管理区域，单击“停止”，即可停止在线服务的部署，避免产生不必要的费用。如果需要继续使用此服务，可单击“启动”恢复。
如果您启用了自动停止功能，服务将在指定时间后自动停止，不再产生费用。

2.7 使用窍门

2.7.1 创建项目时，如何快速创建 OBS 桶及文件夹？

在创建项目时需要选择训练数据路径，本章节将指导您如何在选择训练数据路径时，快速创建OBS桶和OBS文件夹。


1. 在创建自动学习项目页面，单击数据集输入位置右侧的“”按钮，进入“数据集输入位置”对话框。
2. 单击“新建对象存储服务（OBS）桶”，进入创建桶页面，具体请参见《对象存储服务控制台指南》中的**创建桶**章节。

图 2-57 快速创建 OBS 桶



3. 桶创建完成后，选择对应桶名称，单击“新建文件夹”，在“新建文件夹”对话框中，填写文件夹“名称”，单击“确定”完成创建，选择创建的文件夹。
 - 文件夹名称不能包含以下字符：\:*? "<>|。
 - 文件夹名称不能以英文句号 (.) 或斜杠 (/) 开头或结尾。
 - 文件夹的绝对路径总长度不能超过1023字符。
 - 任何单个斜杠 (/) 表示分隔并创建多层级的文件夹。

图 2-58 新建文件夹



2.7.2 增加的数据，如何在自动学习项目中查看？

针对已创建的项目，需要添加新数据时，可以执行如下操作。本章节的操作说明仅针对“物体检测”、“图像分类”、“声音分类”和“文本分类”类型的项目，针对“预测分析”项目，直接在其数据文件中增加数据即可。

获取自动学习项目的数据源

1. 登录ModelArts管理控制台，在左侧菜单栏中选择“自动学习”。
2. 在自动学习项目列表中，您可以查看到项目对应的“数据源”，单击此处链接，可直接跳转至创建项目时选择或者创建的数据集。

📖 说明

针对“预测分析”项目，其数据源指定的是一个OBS路径，并非数据集。其他类型的自动学习项目，其数据源为一个数据集。

图 2-59 查看数据存储路径

项目名称	项目类型	训练状态	数据源	创建时间
exeML-voice	声音分类	部署中 1 训练失败 1	dataset-voice	2020/03/19 15:52:20 GMT+08:00
exeML-bank	预测分析	已完成 1	obs://modelarts-test07/Exeml/d	2020/03/19 15:44:50 GMT+08:00
exeML-yunbao	物体检测	已完成 1	dataset-yunbao	2020/03/18 22:03:19 GMT+08:00
exeML-flowers	图像分类	已完成 1	dataset-flowers	2020/03/18 19:39:04 GMT+08:00
exeML-text	文本分类	部署中 1	dataset-text	2020/03/11 12:09:44 GMT+08:00

将新增数据上传至 OBS

登录OBS管理控制台，访问数据存储路径，将新增数据上传至OBS。

OBS上传文件的操作指导请参见[上传对象](#)。

同步数据至 ModelArts

1. 待数据上传至OBS后，进入ModelArts管理控制台的“自动学习”页面。
2. 在自动学习项目列表中，选择需要增加数据的项目，单击项目名称，进入“数据标注”页面。
3. 在数据标注页面，单击“同步数据源”。

数据同步需要几分钟时间，请耐心等待，待同步数据完成后，新增的数据将被同步至“未标注”或“已标注”页签下。

2.7.3 自动学习项目中，如何进行增量训练？

在自动学习项目中，每训练一次，将自动产生一个训练版本。当前一次的训练结果不满意时（如对训练精度不满意），您可以适当增加高质量的数据，或者增减标签，然后再次进行训练。

说明

- 增量训练目前仅支持“图像分类”、“物体检测”、“声音分类”类型的自动学习项目。
- 为提升训练效果，建议在增量训练时，选择质量较高的数据，提升数据标注的质量。

增量训练的操作步骤

1. 登录ModelArts管理控制台，单击左侧导航栏的自动学习。
2. 在自动学习项目管理页面，单击对应的项目名称，进入此项目的自动学习详情页。
3. 在数据标注页面，单击未标注页签，在此页面中，您可以单击添加图片，或者增删标签。

如果增加了图片，您需要对增加的图片进行重新标注。如果您增删标签，建议对所有的图片进行排查和重新标注。对已标注的数据，也需要检查是否需要增加新的标签。

4. 在图片都标注完成后，单击右上角“开始训练”，在“训练设置”中，在“增量训练版本”中选择之前已完成的训练版本，在此版本基础上进行增量训练。其他参数请根据界面提示填写。

设置完成后，单击“确定”，即进行增量训练。系统将自动跳转至“模型训练”页面，待训练完成后，您可以在此页面中查看训练详情，如“训练精度”、“评估结果”、“训练参数”等。

图 2-60 选择增量训练版本

训练设置

数据集版本名称: V004

训练验证比例: 训练集比例: 0.8, 验证集比例: 0.2

增量训练版本: V001

最大训练时长 (分钟): 60

训练偏好: balance

2.7.4 自动学习生成的模型，存储在哪里？支持哪些其他操作？

模型统一管理

针对自动学习项目，当模型训练完成后，其生成的模型，将自动进入“AI应用管理 > AI应用”页面，如下图所示。模型名称由系统自动命名，前缀与自动学习项目的名称一致，方便辨识。

注意

自动学习生成的模型，不支持下载使用。

图 2-61 自动学习生成的模型

AI应用名称	最新版本	状态	部署类型
huashu@huashu-ml	0.0.1	正常	在线服务
MA-Service-model-11-15-19-54	1.0.0	正常	在线服务/批量服务/边缘服务
model-60447285-flower	0.0.1	正常	在线服务/批量服务/边缘服务

自动学习生成的模型，支持哪些其他操作

- 支持部署为在线服务、批量服务或边缘服务。
在自动学习页面中，仅支持部署为在线服务，如需部署为批量服务或边缘服务，可在“AI应用管理 > AI应用”页面中直接部署。
- 支持发布至市场
将产生的模型发布至AI Gallery，共享给其他用户。

- **支持创建新版本**
创建新版本，仅支持从ModelArts训练作业、OBS、模型模板、或自定义镜像中选择元模型。无法从原自动学习项目中，创建新版本。
- **支持删除模型或其模型版本**

2.7.5 将旧版项目升级到新版

自动学习功能升级为新版，如果您的项目是在旧版中创建，需升级后再使用。未升级的自动学习项目，无法进行数据标注、训练以及部署等操作。

📖 说明

“预测分析”类型的项目，可以不执行升级，直接使用新版自动学习。

升级到新版

1. 登录ModelArts管理控制台，在左侧菜单栏中选择“自动学习”，进入“自动学习”项目列表页面。
 2. 查找您的旧版项目。在“自动学习”列表中，如果项目是旧版的，在项目名称处带有标识，针对此类项目，单击操作列的“升级”。
- 如果您的项目是新版的，则操作列中不存在“升级”按钮。

图 2-62 查找旧版项目

⚠️ 温馨提示：旧版自动学习项目请升级到新版后再使用。

项目名称	项目类型	训练状态	数据源	创建时间	描述	操作
exeML-84d6	图像分类	模型发布失败 2	dataset-Special	2019/12/18 16:33:56 GMT+08:00	-	删除
exeML-c737	文本分类	已完成 4 模型发布失败 2	dataset-eaf2	2019/12/18 16:04:31 GMT+08:00	-	删除
exeML-70f1	声音分类	已完成 4	obs://getobs/voice/	2019/12/05 09:43:53 GMT+08:00	this is a project migrated fro...	升级 删除
exeML-a96b	预测分析	已完成 5 训练失败 1	obs://getobs/datasets/data/tra	2019/12/05 09:43:31 GMT+08:00	this is a project migrated fro...	删除

3. 在弹出的对话框中，设置需保存的“数据集名称”，以及数据集“存储路径”，然后单击“确定”启动升级。

图 2-63 升级项目

升级项目exeML-70f1

* 数据集名称

* 存储路径

⚠️ 请选择不在 /getobs/voice/ 目录下的OBS路径。

等待项目升级，大概几分钟之后，项目将被升级为新版，项目名称处不再有旧版标识，且操作列不再存在“升级”按钮。