

机器学习PAI

产品简介

产品简介

什么是机器学习PAI

机器学习指机器通过统计学算法，对大量的历史数据进行学习从而生成经验模型，利用经验模型指导业务。目前机器学习主要在以下方面发挥作用：

- 营销类场景：商品推荐、用户群体画像、广告精准投放
- 金融类场景：贷款发放预测、金融风险控制、股票走势预测、黄金价格预测
- SNS关系挖掘：微博粉丝领袖分析、社交关系链分析
- 文本类场景：新闻分类、关键词提取、文章摘要、文本内容分析
- 非结构化数据处理场景：图片分类、图片文本内容提取OCR
- 其它各类预测场景：降雨预测、足球比赛结果预测

笼统地讲，机器学习可以分为三类：

- 有监督学习（Supervised Learning）：指每个样本都有对应的期望值，通过模型搭建，完成从输入的特征向量到目标值的映射。典型的案例就是回归和分类问题。
- 无监督学习（Unsupervised Learning）：指在所有的样本中没有任何目标值，期望从数据本身发现一些潜在的规律，例如一些简单的聚类。
- 增强学习（Reinforcement Learning）：相对来说比较复杂，是指一个系统和外界环境不断地交互，获得外界反馈，然后决定自身的行为，达到长期目标的最优化。其中典型的案例就是阿法狗下围棋，或者无人驾驶。

阿里云机器学习PAI平台是构建在阿里云MaxCompute（原ODPS）计算平台之上，集数据处理、建模、离线预测、在线预测为一体的机器学习平台。阿里云机器学习封装了阿里巴巴集团内成熟的算法，向机器学习用户提供了更简易的操作体验，真正做到让人工智能触手可及。

产品架构

概述

阿里云机器学习PAI平台的产品架构如下图所示。



说明如下：

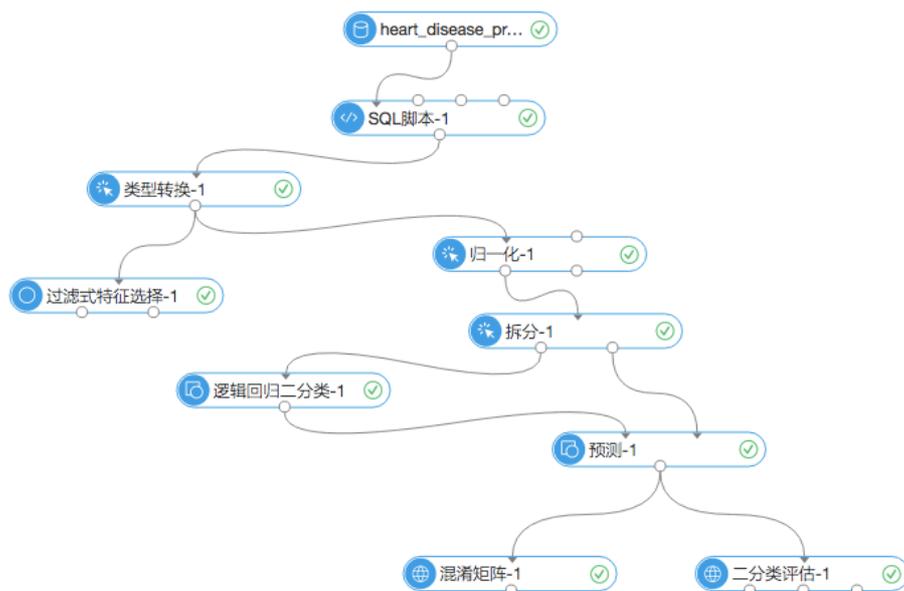
- 基础设施层：CPU计算集群。
- 计算框架层：包括MapReduce、SQL、MPI等计算方式，分布式计算架构主要执行并行化计算分发任务。
- 核心产品功能层：提供PAI-AutoLearning、PAI-Studio、PAI-DSW三种建模方式，从左到右，建模的灵活度更高。从右到左，建模的技术要求降低。其中Studio中包括了数据预处理、特征工程、机器学习算法、深度学习等基本组件。所有算法组件全部脱胎于阿里巴巴集团内部成熟的算法体系，经受过PB级别业务数据的锤炼。此外，在模型建模基础上，提供模型在线服务一键部署功能。最后，用户还可以通过在智能生态市场快速获取业务解决方案或模型算法，进行相关业务与技术的高效对接。
- 业务应用层：阿里巴巴内部的搜索系统、推荐系统、蚂蚁金服等项目在进行数据挖掘工作时，都是依赖机器学习平台产品。机器学习平台的业务场景包含了金融、医疗、教育、交通、安全等各个领域。

产品优势

阿里云机器学习平台的产品主要优势如下。

良好的交互设计

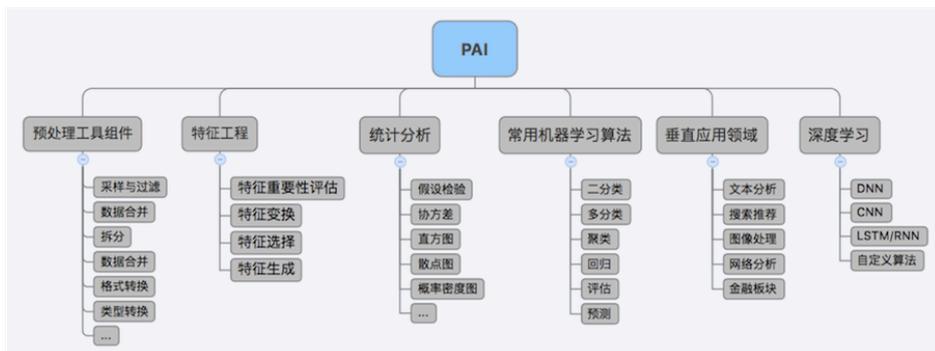
通过对底层的分布式算法封装，提供拖拉拽的可视化操作环境，让数据挖掘的创建过程像搭积木一样简单。缩短了您与数据的距离，真正实现了数据的触手可及。同时也提供了命令行工具，方便您将算法嵌入到自己的工



程中。

优质、丰富的机器学习算法

机器学习平台上的算法都是经过阿里大规模业务锤炼而成的。从算法的丰富性角度来看，阿里云机器学习平台不仅提供了基础的聚类、回归类等机器学习算法，也提供了文本分析、特征处理等比较复杂的算法。



与阿里产品完美配合

使用阿里云机器学习平台计算的模型直接存储在MaxCompute（原ODPS）上，可以配合其它阿里云的产品组件加以利用。



一站式的机器学习体验

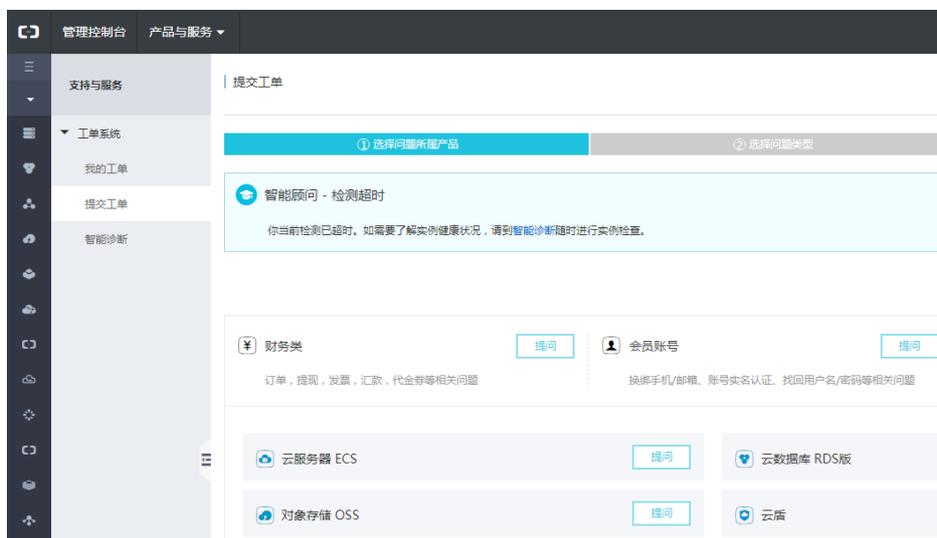
阿里云机器学习平台除了提供模型训练功能，还提供在线预测以及离线调度功能，让机器学习训练结果和业务可以无缝衔接。

支持主流深度学习框架

阿里云机器学习平台已经包含了Tensorflow、Caffe、MXNet这三款主流的机器学习框架，底层提供M40型号的GPU卡进行训练。

优质的技术保障

阿里云机器学习算法平台的背后是阿里巴巴的算法科学家和阿里云的技术保障团队，在使用过程中遇到任何问题都可以到工单系统提交工单或者直接与相关接口人联系，具体请参见用户交流。



PAI权限管理文档

PAI成员权限管理文档

PAI分为PAI-STUDIO、PAI-DSW、PAI-EAS 3款子产品。其中PAI-STUDIO为拖拽式的模型训练平台，PAI-DSW为Notebook交互式模型训练平台，PAI-EAS为模型在线服务平台。3款产品各有各的权限点。具体权限点请分别参考下方目录：

- STUDIO权限管理说明
- DSW权限管理说明
- EAS权限管理说明

功能特性

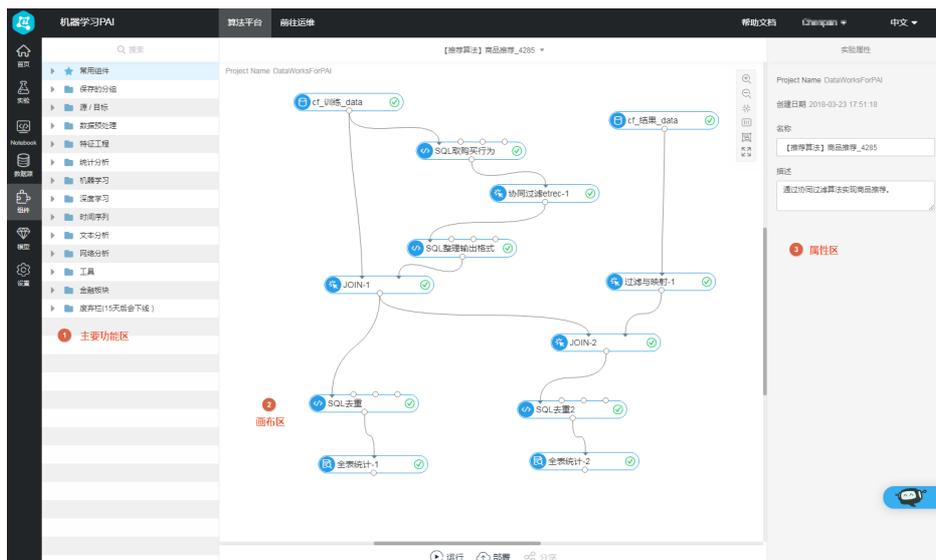
阿里云机器学习平台是基于阿里云的云计算平台，具有超大规模的数据处理能力和分布式的存储能力，同时整个模型支持超大规模的建模以及计算。

该平台主要分为三层：

- 第一层：Web UI界面
- 第二层：机器学习算法层
- 第三层：MAXCOMPUTE平台层

Web UI界面

阿里云机器学习平台的Web UI界面如下图所示：



主要区域

Web UI界面主要由如下区域组成：

- 主要功能区：展示了各组件名称。
- 画布区：使用者可以用鼠标将相应的组件拖拽到画布上，形成一个有向的工作流，完成从元数据到数据处理再到建模等一系列的数据挖掘工作。
- 属性区：该区域可以设置组件内参数的信息。

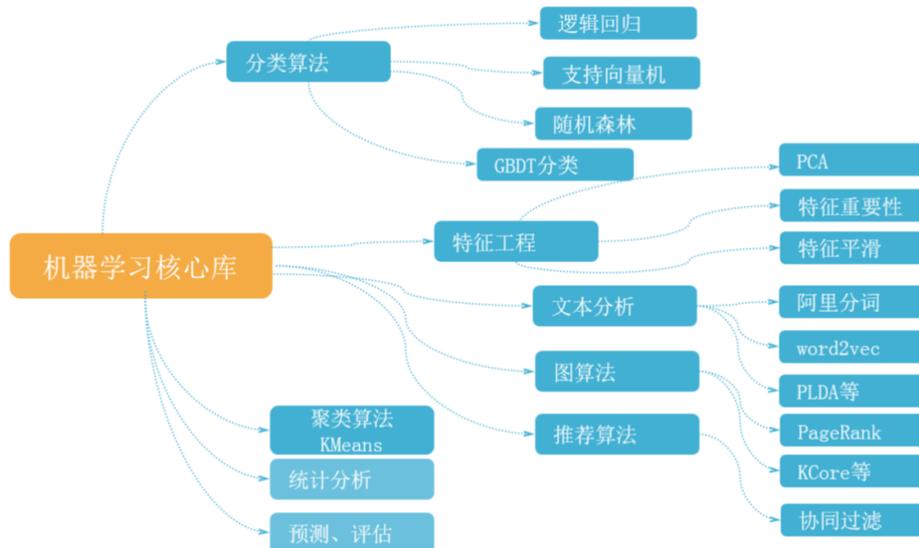
主要功能

Web UI界面主要功能如下：

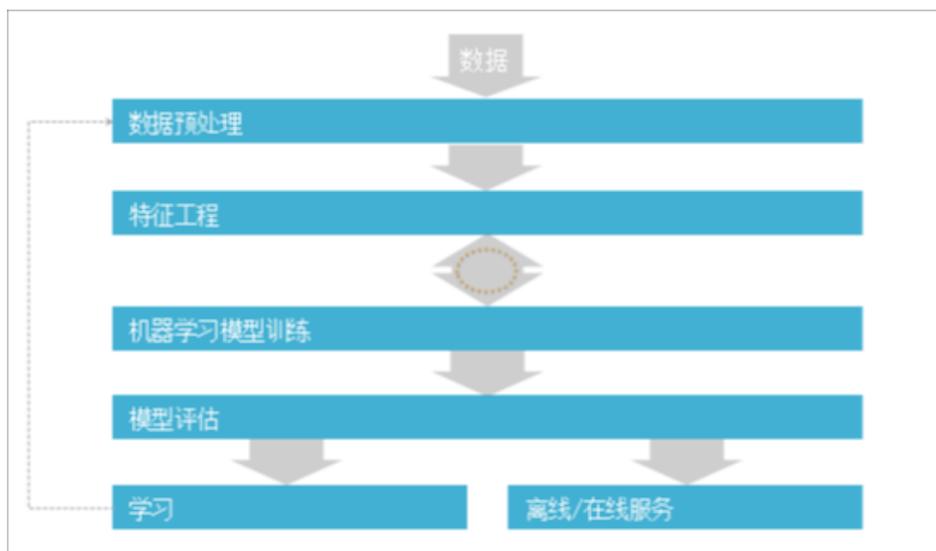
- 搜索功能：当有很多数据、表、实验时，可以通过搜索功能快速查找所需的资料。
- 实验列表：通过单击实验名称，在画布上显示该实验的有向实验流图，可以继续之前没有完成的实验。
- 数据表：类似于文件管理器，可以查看所有数据表。
- 算法和工具列表：包含了常用的机器学习算法组件等核心功能。
- 模型列表：通过该功能，使用者可以管理所有的模型。

算法平台

阿里云机器学习平台拥有完整的算法库，如下图所示：



通常情况下，一个数据智能应用的完整链路很长，从数据的ETL到预处理、特征工程、模型训练、评估和服务。以前经常不得不在多个工具中来回穿梭才能完成任务，而机器学习平台致力于打造一站式的开发环境，提供了从元数据到模型部署整套流程。通过基本组件，可以搭建各个垂直场景下的解决方案，节省了大量切换环境的损耗。一个典型的算法开发流程如下：

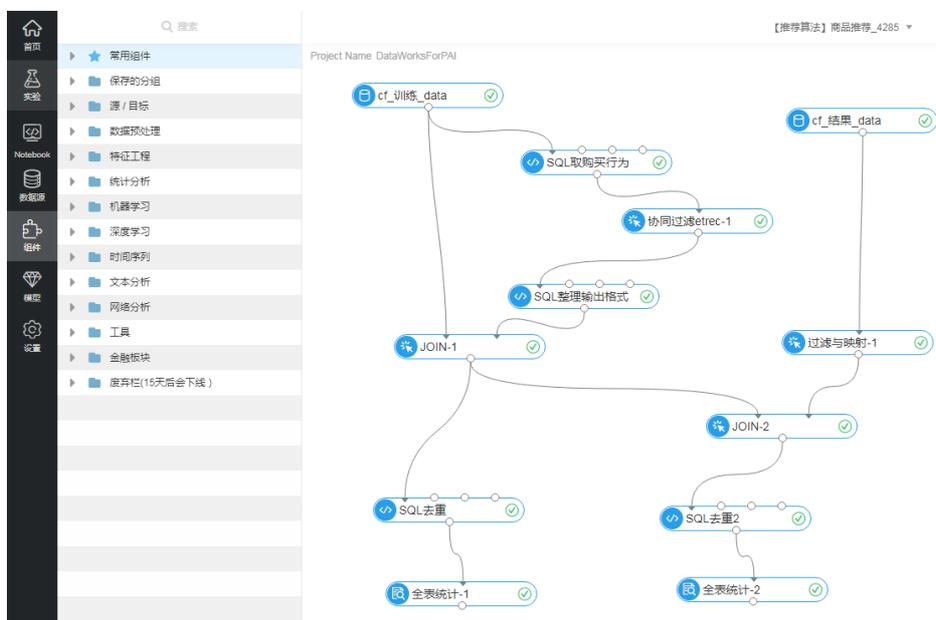


完善的数据挖掘组件

提供整套的数据挖掘组件，包括数据的预处理、特征抽取、模型训练、预测和评估。将数据导入平台后，可通过拖拉的方式，借助实验组件灵活地拼装自己的试验流程来解决自身的业务场景。

可视化建模

操作界面通过拖拉拼接实验，提供了可视化建模的功能，方便您的操作。从左边的组件框中拖拉组件到右边的实验区即可进行实验的搭建，如下图所示。



数据可视化

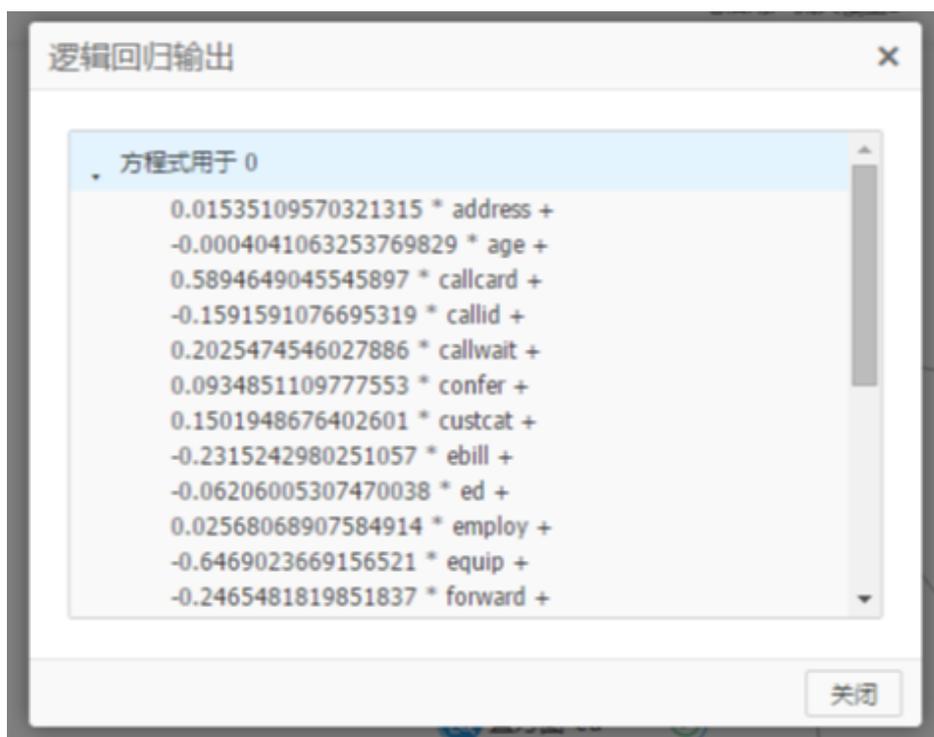
对于每个输出型组件，都可以通过右键单击组件来查看可视化输出模型。可视化输出有多种表示方法，包括折

线图、点图和柱形图，如下图所示。



模型可视化

模型的可视化管理：在左侧的菜单中选择**模型**页签，打开**实验模型**文件夹，可查看所有实验的模型列表。右键单击模型，选择**查看模型**，即可查看对应的模型结果，如下图所示。



算法组件

阿里云机器学习平台包含数据预处理、特征工程、统计分析、机器学习、深度学习框架、预测与评估这一整套的机器学习算法组件，共100余种。同时包含了Tensorflow、Caffe、MXNet这三款主流的机器学习框架，底

层提供M40型号的GPU卡进行训练。

用户交流

阿里云机器学习平台以社区化的方式进行用户运营，会经常举行用户活动，同时为企业级用户提供高效支持。如果您在产品使用过程中遇到问题或bug，请通过工单系统和机器学习平台内部的问答机器人解决。用户交流群主要用来进行机器学习算法心得交流以及相关活动的推广。

加入阿里云机器学习钉钉群，群二维码如下：



入群链接：[邀请链接](#)

另外也可以关注我们在云栖社区的官方公众号，会有精彩文章推送。

相关地址

[产品入口页](#)

[收费说明](#)

[算法组件文档](#)

深度学习文档

在线预测

离线调度

案例说明

产品BUG反馈，工单系统