

阿里云机器学习

产品简介

产品简介

什么是阿里云机器学习

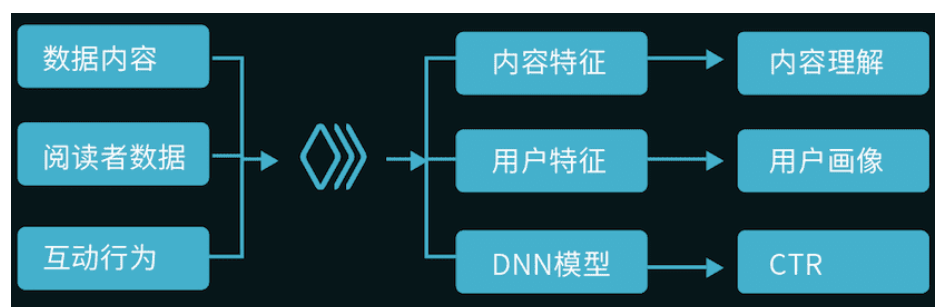
概述

机器学习指通过统计学算法，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。

阿里云机器学习平台（Machine Learning Platform for AI），为用户提供低门槛、高性能的一站式机器学习服务。用户基于机器学习平台可以快速搭建机器学习实验并且将进行部署，使得算法与业务无缝结合。阿里云机器学习平台封装了阿里巴巴集团内成熟的算法，目前已经对外服务数万开发者及企业用户。用户可以基于机器学习平台快速搭建商品推荐、金融风控、图像识别、语音识别等人工智能服务，真正做到让人工智能触手可及。

案例-新浪微博

新浪微博是中国领先的社交媒体平台，拥有1.65亿日活跃用户（DAU），3.76亿月活跃用户（MAU），在月活跃用户的移动端占比高达92%。涉及到的数据化运营场景包含但不限于：



挑战

用户每日都在平台产生大量的数据，数据经过处理达到百亿特征、千亿样本级别，如何高效地对如此大规模的数据进行有效的计算和分析，给底层的计算引擎带来极大挑战。



解决方案及架构

机器学习平台的鲲鹏分布式计算引擎作为微博机器学习平台的一部分，为超大规模机器学习计算场景提供了有效的解决方案。

产品架构

阿里云机器学习平台的基础架构如下图所示。



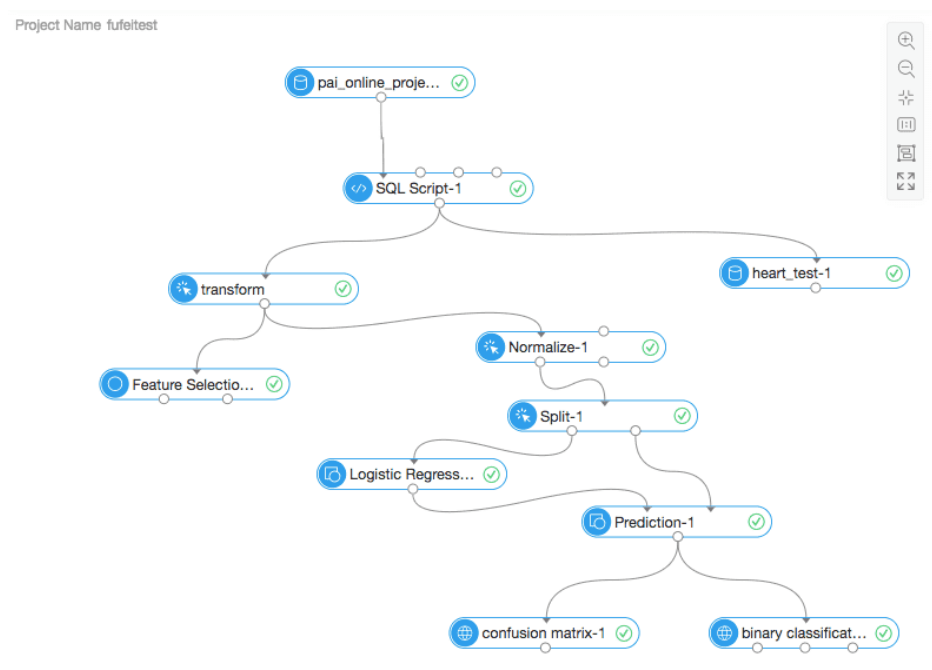
说明如下：

- 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。
- 模型算法层：包含数据预处理、特征工程、机器学习算法等基本组件，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。
- 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。

功能概述

用户操作界面

通过拖拉拽的方式搭建机器学习实验，实现零门槛上手。



可视化展示机器学习全流程计算结果。



丰富的算法组件

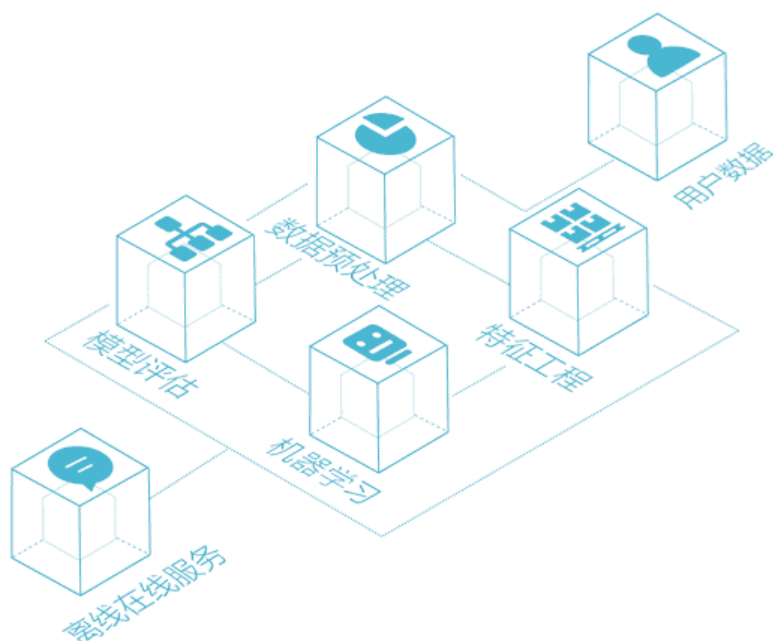
提供回归、分类、聚类、文本分析、关系挖掘等100余种算法组件。

包括预处理工具软件、特征工程、分析系统、垂直应用领域、常用机器学习算法、金融板块算法等。



一站式服务

提供一站式的服务体验，数据清洗、特征工程、机器学习算法、评估都可以在机器学习平台上一站式使用。



超大规模计算框架

通过异步、并行通信等方面的大量优化，支持高吞吐低延迟的参数交换、多种流式IO以及Exactly Once的Failover，从而支撑了千亿特征、万亿参数模型规模的训练任务。在对大量算法业务的长期支持下，积累了从线性模型到决策树再到深度稀疏模型等完整的算法体系。在大规模稀疏特征场景下，支持动态特征调控，能够在任意时刻增删模型特征，是推荐、广告、搜索等CTR预估场景下的算法利器。