

主界面功能模块介绍

- 数据挖掘入口
- 管理模块
 - 实验管理
 - ‘实验’ 页签
 - ‘案例’ 页签
 - 模型管理

数据挖掘入口

入口一：产品右侧导航栏数据挖掘模块。

入口二：系统界面的各个模块快捷菜单。

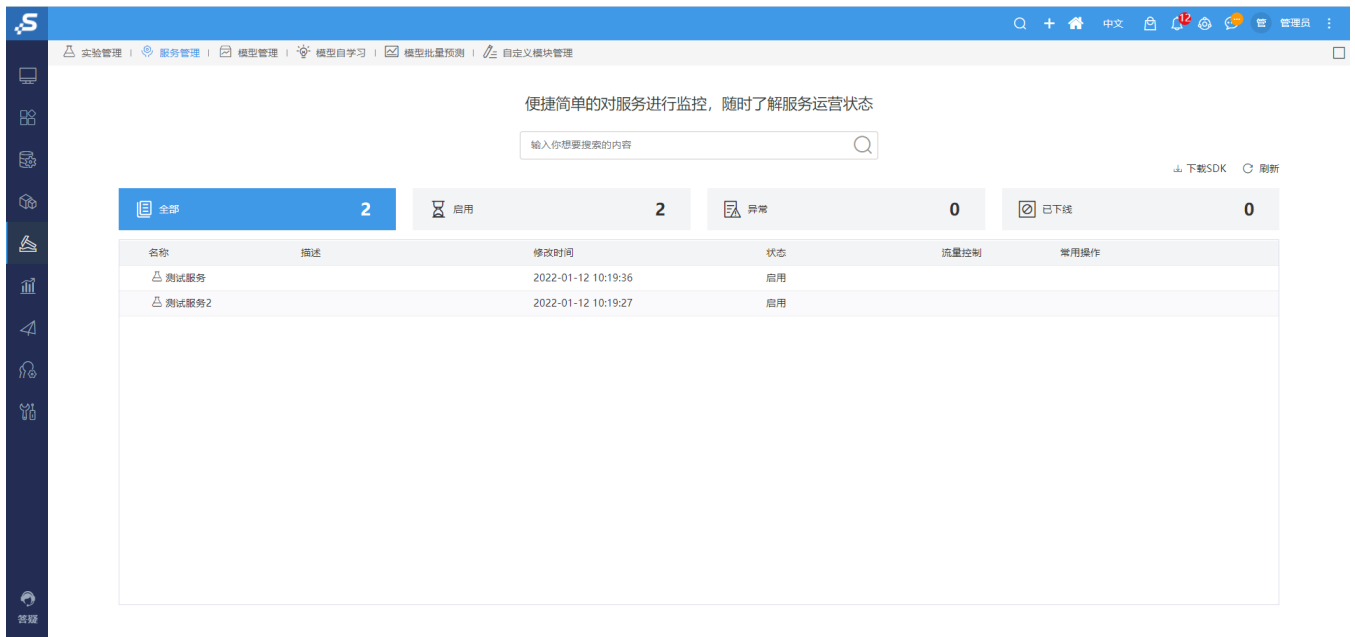


管理模块

从入口进来后，默认进入‘实验管理界面’，在该页面的顶部是数据挖掘各管理模块的入口，具体页面展示了相关内容的基本信息以及常用操作按钮，而各模块主界面是一个表格树目录展示内容。部分模块对内容进行分类，在表格树目录上方有不同页签，单击页签可进入不同页面，如下图：



其余运维管理界面大同小异，用页签和表格树目录来展示各项基本信息，以及一些常用操作按钮，下图为服务管理页面：



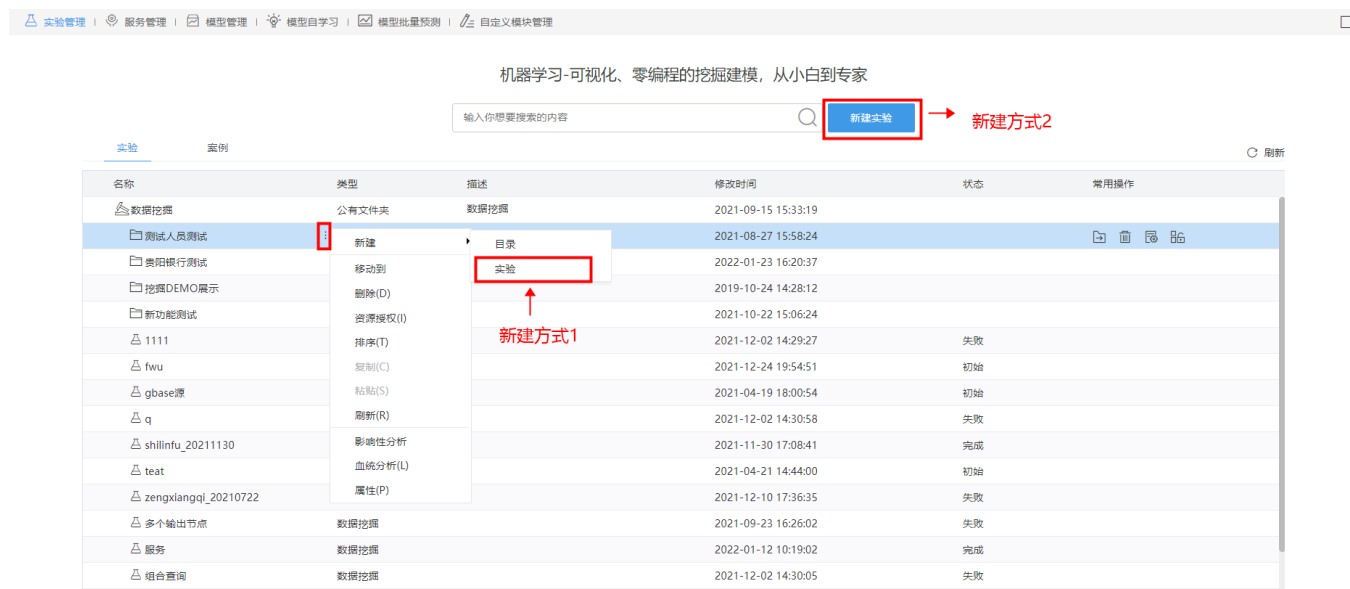
其余模型自学习，模型批量预测，自定义模块管理基本相同。

实验管理

由于‘实验管理’界面的功能复杂于其他运维管理模块，下面以‘实验管理’模块进行简单介绍：

‘实验’页签

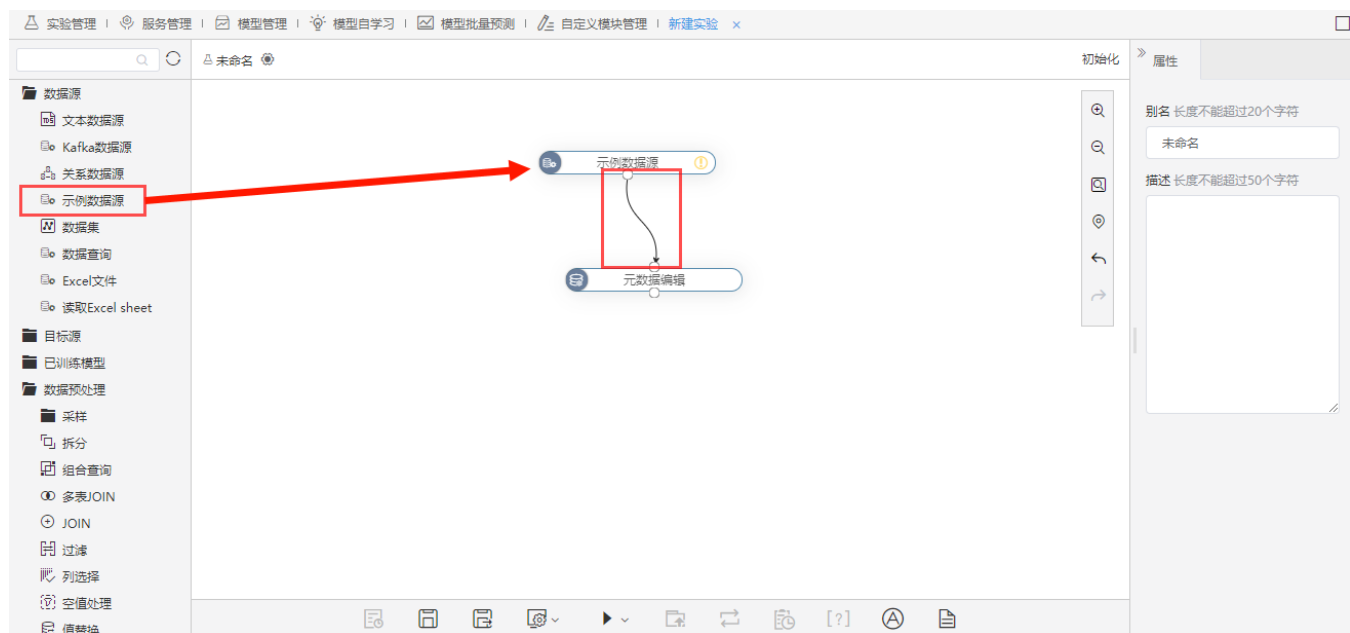
该页签展示保存在‘数据挖掘’，‘公共空间’，‘我的空间’文件目录下的数据挖掘资源。其中，‘新建’实验的入口有两处，如下图：



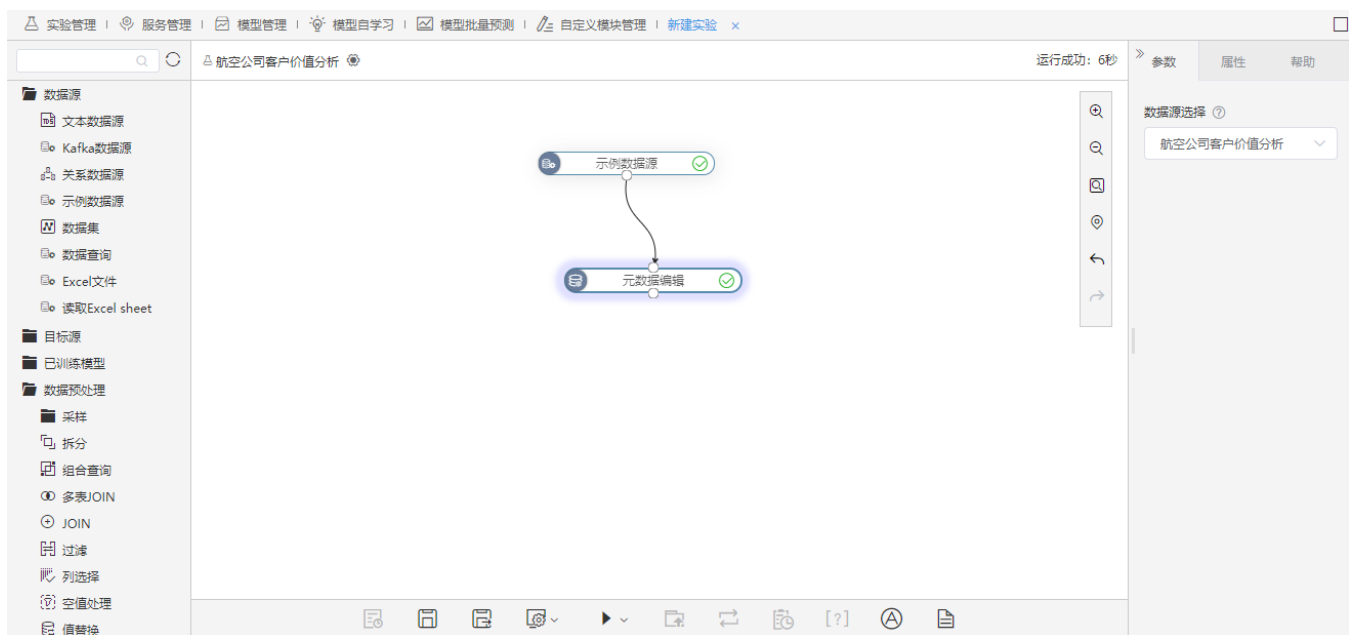
新建方式1: 在文件夹右键‘新建’中单击‘实验’;

新建方式2: 单击‘新建实验’。

挖掘实验基于空白画布，根据需要将相应节点拖拽至画布区域，并按照数据处理流程在节点间进行连线，如下图：



挖掘实验就是由一个个节点以及它们之间的连线构成，按照需要进行实验构建，保存后，就可以获得专属的数据挖掘模型：



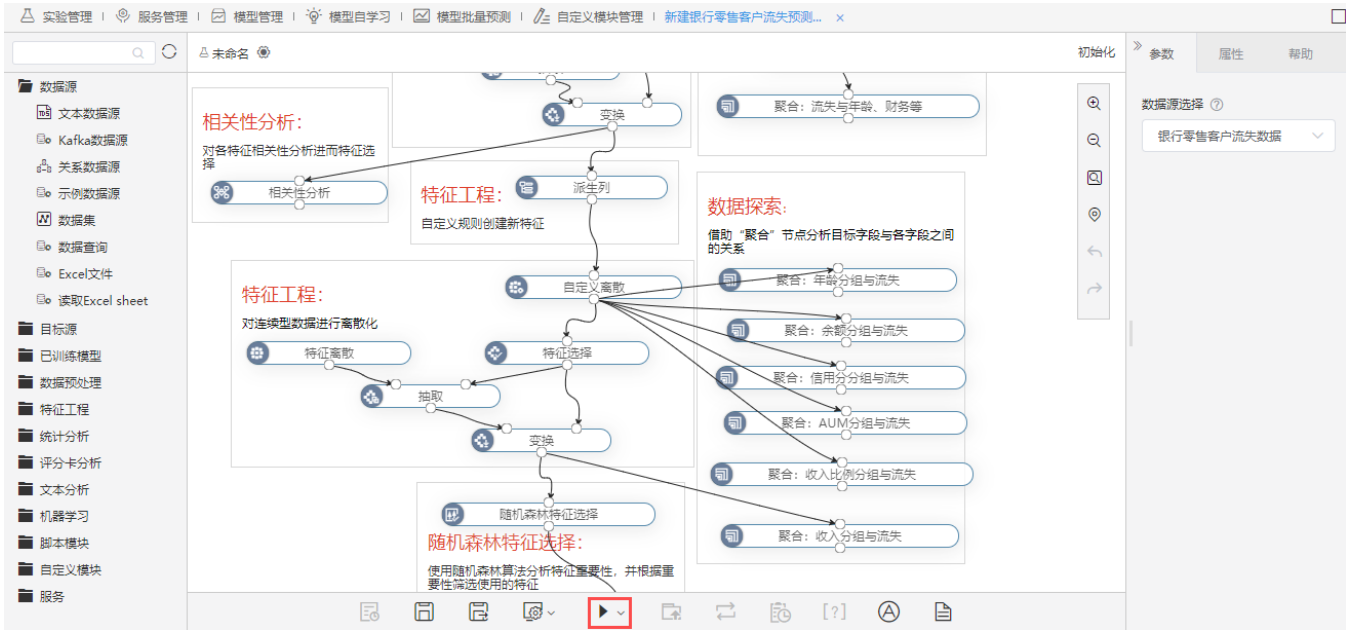
快来动手搭建属于自己的挖掘实验吧~

‘案例’ 页签

该页面展示了产品内置的案例，单击相应案例即可以进入实验界面。



画布区域已经提供搭建好的挖掘实验流程，可以直接运行：



注：部分案例的对应说明，仅供学习参考。

模型管理

用于管理一些已训练好的模型。

实验管理 服务管理 模型管理 模型自学习 模型批量预测 自定义模块管理 新建银行零售客户流失预测...			
统一管理机器学习实验训练生成的模型			
输入你想要搜索的内容			
上传模型 刷新			
名称	描述	修改时间	常用操作
模型提升回归树		2021-11-05 14:49:01	
模型提升决策树测试		2021-11-05 14:48:05	
朴素贝叶斯测试		2021-11-05 14:47:17	
支持向量机测试		2021-11-05 14:46:21	
多模型和测试		2021-11-05 14:40:33	导出PMML 导出模型
cp-model		2021-11-02 18:08:42	
GBDT唐		2021-10-27 14:28:54	
yjl_1026		2021-10-26 16:25:27	
save_model		2021-10-22 12:31:14	
模型自学习测试模型1		2021-10-20 11:11:05	
aaa		2021-10-09 18:47:10	
线性回归0927		2021-09-27 16:13:20	
local+webhdfs+0927		2021-09-27 10:03:17	

支持的操作包括：

操作	说明
上传模型	上传导出模型生成的ZIP文件，并生成对应的模型记录，可在实验界面左侧资源树查看生成的节点。
刷新	刷新表格。
删除	删除选择的模型。
导出模型	用于将训练得到的模型导出，并在产品的其他服务器使用。 比如在一台服务器导出的模型文件，可通过“上传模型”功能导入到其他服务器使用。

导出
PMML
模型

用于将训练得到的模型转化为PMML模型文件（通用）。用户可将PMML模型文件载入Python或其他平台（如KNIME）中进行预测，预测结果与产品预测结果一致。

支持“导出PMML模型”功能的算法有：梯度提升决策树、线性回归、逻辑回归、支持向量机、多层感知机、朴素贝叶斯、随机森林、K均值。其中，梯度决策树、线性回归、多层感知机、朴素贝叶斯的PMML 文件输出预测概率，支持向量机、随机森林的PMML文件输出预测标签。



- 1) 暂不支持向量特征生成PMML模型文件。
- 2) 目前导出PMML模型功能只支持部分算法，特征工程和数据处理部分暂不支持，因此用户导出数据模型到其它平台后，需要进行数据处理和特征工程。



PMML

PMML是数据挖掘的一种通用的规范，它用统一的XML格式来描述我们生成的机器学习模型，可以将其转化为标准的XML格式来存储。当需要将这个PMML的模型用于部署的时候，可以使用目标环境的解析PMML模型的库来加载模型，并做预测。

可以看出，要使用PMML需要两步的工作，第一块是将离线训练得到的模型转化为PMML模型文件，第二块是将PMML模型文件载入在线预测环境进行预测。这两块都需要相关的库支持。