

BONC 东方国信

内部资料

数慧空间

产品操作手册

(前台)

编写	刘昊	编写 时间	2018-04-26
审核		审核 时间	
文档版本	V2.0		

东方国信版权所有

文档修订摘要

日期	版本号	修订章节	描述	作者	评审者	评审日期
2017-11-02	V1.0					
2018-05-10	V2.0			刘昊		

目 录

第 1 章	阅读指南	1
1.1	目的	1
1.2	使用建议	1
1.3	阅读对象	1
1.4	名词解释	1
1.5	参考资料	1
第 2 章	系统简介	2
2.1	系统功能和用途	2
2.2	系统运行环境说明	2
第 3 章	数慧空间操作指南	3
3.1	系统登录	3
3.1.1	业务描述	3
3.1.2	功能界面	4
3.1.3	操作步骤	4
3.2	首页	4
3.2.1	业务描述	4
3.2.2	功能界面	5
3.2.3	操作步骤	5
3.3	模型管理	6
3.3.1	模型列表	6
3.3.1.1	业务描述	6
3.3.1.2	功能界面	6
3.3.1.3	操作步骤	6
3.3.1.4	注意事项	7
3.3.2	模型创建	7
3.3.2.1	业务描述	7
3.3.2.2	功能界面	7
3.3.2.3	操作步骤	7
3.3.2.4	注意事项	8
3.3.3	模型版本管理	8
3.3.3.1	业务描述	8
3.3.3.2	功能界面	8
3.3.3.3	操作步骤	8
3.3.3.4	注意事项	9
3.3.4	模型开发	9
3.3.4.1	数据源管理	9
3.3.4.1.1	业务描述	9
3.3.4.1.2	功能界面	9
3.3.4.1.3	操作步骤	10
3.3.4.1.4	注意事项	10
3.3.4.2	数据集管理	10
3.3.4.2.1	业务描述	10
3.3.4.2.2	功能界面	11
3.3.4.2.3	操作步骤	11

3.3.4.2.4	注意事项	12
3.3.4.3	离线模型日志管理	12
3.3.4.3.1	业务描述	12
3.3.4.3.2	功能界面	12
3.3.4.3.3	操作步骤	12
3.3.4.3.4	注意事项	12
3.3.4.4	模型开发-离线方式	12
3.3.4.4.1	业务描述	12
3.3.4.4.2	功能界面	13
3.3.4.4.3	操作步骤	14
3.3.4.4.4	注意事项	14
3.3.4.5	模型开发-在线	14
3.3.4.5.1	业务描述	14
3.3.4.5.2	功能界面	16
3.3.4.5.3	操作步骤	16
3.3.4.5.4	注意事项	24
3.3.5	模型详情	24
3.3.5.1	业务描述	24
3.3.5.2	功能界面	25
3.3.5.3	操作步骤	25
3.3.5.4	注意事项	25
3.3.6	模型上下线	26
3.3.6.1	业务描述	26
3.3.6.2	功能界面	26
3.3.6.3	操作步骤	26
3.3.6.4	注意事项	27
3.3.7	模型审核及授权	27
3.3.7.1	业务描述	27
3.3.7.2	功能界面	27
3.3.7.3	操作步骤	27
3.3.7.4	注意事项	28
3.3.8	模型调度	28
3.3.8.1	业务描述	28
3.3.8.2	功能界面	28
3.3.8.3	操作步骤	28
3.3.8.4	注意事项	29
3.3.9	模型评估	29
3.3.9.1	业务描述	29
3.3.9.2	功能界面	30
3.3.9.3	操作步骤	30
3.3.9.4	注意事项	30
3.4	系统管理	31
3.4.1	客户端监控	31
3.4.1.1	业务描述	31
3.4.1.2	功能界面	31
3.4.1.3	操作步骤	31

3.4.1.4	注意事项	31
3.4.2	客户端管理	32
3.4.2.1	业务描述	32
3.4.2.2	功能界面	32
3.4.2.3	操作步骤	32
3.4.2.4	注意事项	32
3.4.3	模型类别管理	32
3.4.3.1	业务描述	32
3.4.3.2	功能界面	33
3.4.3.3	操作步骤	33
3.4.3.4	注意事项	33
3.4.4	企业账号管理	33
3.4.4.1	业务描述	33
3.4.4.2	功能界面	34
3.4.4.3	操作步骤	34
3.4.4.4	注意事项	34
3.4.5	用户管理	34
3.4.5.1	业务描述	34
3.4.5.2	功能界面	35
3.4.5.3	操作步骤	35
3.4.5.4	注意事项	35
3.5	流程开发	35
3.5.1	预测模型管理	35
3.5.1.1	业务描述	35
3.5.1.2	功能界面	36
3.5.1.3	操作步骤	36
3.5.1.4	注意事项	36
3.5.2	样例库管理	36
3.5.2.1	业务描述	36
3.5.2.2	功能界面	37
3.5.2.3	操作步骤	37
3.5.2.4	注意事项	37
3.5.3	缓存配置	37
3.5.3.1	业务描述	37
3.5.3.2	功能界面	38
3.5.3.3	操作步骤	38
3.5.3.4	注意事项	38
3.6	权限说明	39
3.6.1.1	业务描述	39
3.6.1.2	功能界面	39
3.6.1.3	注意事项	39

第1章 阅读指南

1.1 目的

本手册为数慧空间操作手册，通过阅读该手册，可以帮助用户了解数慧空间系统的各项功能的使用。也可以帮助业务使用人员了解系统的功能，并根据每个功能的描述指导操作对应的功能。

注意：本手册截图中的数据为测试数据。

1.2 使用建议

阅读本手册，需要了解本系统的主要功能、以及功能要点，再根据想要完成的功能查看目录，查找具体的细节操作。使用时，确定自己的账号是否有相应的权限，否则，将无法正常使用。

用户在使用时，如果时间充裕，可以通读文档一遍，因为系统中有很多功能是相互关联，相互协调的；文档中的操作步骤描述，皆是通过正常的操作步骤执行。当然，您如果想快速上手，只想知道使用系统中的功能，则可直接看目录查找相应的功能，根据操作步骤依次执行即可。您也可以完成相应功能的操作。

1.3 阅读对象

本手册针对数慧空间的使用用户，主要包括：

- ◆ 模型管理人员
- ◆ 业务使用人员
- ◆ 模型开发人员

1.4 名词解释

无

1.5 参考资料

第2章 系统简介

2.1 系统功能和用途

本系统的名称：数慧空间

开发单位：北京东方国信科技股份有限公司

东方国信的数慧空间，是基于大数据分析的可视化数据挖掘及模型管理平台，实现一站式的线上数据挖掘及模型闭环管理，提供全程自主配置化的模型开发能力，降低模型实现的门槛，提高模型建设的效率，提升模型的价值和复用性。

面向以下三类人员提供服务：

- 面向业务人员：支持不同行业、不同场景、不同业务目标的数据挖掘需求，通过算法组件拖拽、作业引导、可视化数据探索的方式，使业务人员轻松、简便的创建模型。
- 面向技术人员：产品各个算法组件支持详细的参数调整，并对算法应用后的挖掘结果通过专业指标进行评估，便于对模型进行调优，以进行更专业、更有效的数据挖掘。
- 面向管理人员：平台所提供的技术管理手段，帮助模型管理人员从模型状态、模型信息上都能实现更全面的管理，有效地跟踪模型完整的生命周期，以及沉淀模型知识库。

2.2 系统运行环境说明

单机服务器版：

操作系统				
Linux-centos6.5				
硬件环境				
内存	由于 python 脚本在内存里处理相当量的数据，需要一台不少于 64G 内存的机器。	磁盘	用于存储上传的数据文件、python 脚本，磁盘空间需不少于 1T。	
软件环境				

1. jdk1.8	2. tomcat7 开放 tomcat 相应端口供外部访问	3. mysql5.6.33 及 5.7 新增用户并授权该用户远程连接、创建库、创建表等权限	4. python3.5	
浏览器				
最佳：谷歌浏览器。	其次：IE10 以上版本浏览器、火狐浏览器。	其他浏览器暂不建议使用。		

分布式云版：

操作系统				
Linux-centos6.5				
硬件环境				
Web 服务器：	磁盘 1T 以上	内存不少于 16G		
计算节点：	Spark 集群	Spark-2.2.0	至少 3 台 每台内存至少 32G	
软件环境				
1. jdk1.8	2. tomcat7 开放 tomcat 相应端口供外部访问	3. mysql5.6.33 新增用户并授权该用户远程连接、创建库、创建表等权限。	4. python3.5	
浏览器				
最佳：谷歌浏览器。	其次：IE10 以上版本浏览器、火狐浏览器。	其他浏览器暂不建议使用。		

第3章 数慧空间操作指南

3.1 系统登录

3.1.1 业务描述

登录并使用数慧空间的途径。

3.1.2 功能界面



3.1.3 操作步骤

用户使用浏览器输入网址，打开数慧空间，进入登录页面。输入用户名、密码及验证码，点击登录，进入数慧空间首页。

登录后可在模型列表，进行注销。

3.2 首页

3.2.1 业务描述

用户成功登录后，进入数慧空间首页。

首页上方区域，展示模型的统计信息，包括模型总数、访问次数、模型应用效果信息。

首页右方区域，以模型热度统计来展示模型的使用情况。

首页中间区域，是数慧空间的核心能力展示，也是从模型创建到开发到模型上下线再到审核及评估的完整闭环的能力入口。

3.2.2功能界面



3.2.3操作步骤

- 用户点击模型列表——进入“模型管理”首页面；
- 用户点击模型创建——进入“模型创建”页面；
- 用户点击模型创建——进入“模型类别管理”页面；
- 用户点击数据源管理——进入“数据源管理”页面；
- 用户点击数据集管理——进入“数据集管理”页面；
- 用户点击数据集管理——进入“可视化挖掘工具”页面；
- 用户点击模型上下线——进入“模型上线”首页面；
- 用户点击模型上下线——进入“模型下线”首页面；
- 用户点击上线审核——进入“上线审核”页面；
- 用户点击模型授权——进入“模型授权”页面；
- 用户点击模型使用统计——进入“调度管理”页面；
- 用户点击模型效果评估——进入“调度监控”页面。

用户点击模型效果评估——进入“模型使用统计”页面。

用户点击模型效果评估——进入“模型效果评估”页面。

3.3 模型管理

3.3.1 模型列表

3.3.1.1 业务描述

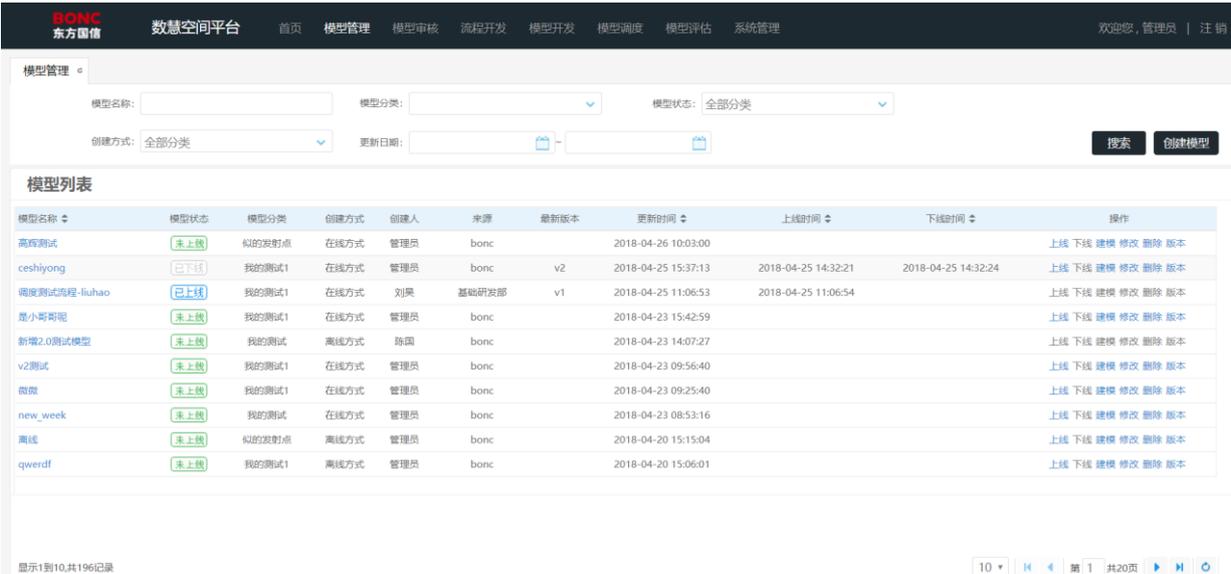
对应菜单栏——首页——模型管理。

页面上方区域支持模型搜索。

页面中间区域，以列表展示所创建的模型，包括模型名称、模型状态、模型分类、创建方式、更新日期。

根据不同系统访问人员的权限不同，所看到的列表内容有所不同：管理员可查看全部，普通用户只可查看自己创建和别人分享的模型。

3.3.1.2 功能界面



The screenshot displays the 'Model Management' (模型管理) interface. At the top, there is a navigation bar with the BONC logo and various menu items like '首页', '模型管理', '模型审核', etc. Below the navigation bar, there is a search and filter section with input fields for '模型名称', '模型分类', '模型状态', and '创建方式', along with a date range selector for '更新日期'. A '搜索' (Search) button and a '创建模型' (Create Model) button are also present. The main content area is titled '模型列表' (Model List) and contains a table with the following columns: 模型名称 (Model Name), 模型状态 (Model Status), 模型分类 (Model Category), 创建方式 (Creation Method), 创建人 (Creator), 来源 (Source), 最新版本 (Latest Version), 更新时间 (Update Time), 上线时间 (Online Time), 下线时间 (Offline Time), and 操作 (Operations). The table lists several models, including '高辉测试', 'ceshiyong', '调度测试流程-liuhao', '最小哥斯基', '新增2.0测试模型', 'v2测试', '微微', 'new_week', '离线', and 'qwerdf'. Each row includes a '操作' column with links for '上线', '下线', '建模', '修改', '删除', and '版本'. At the bottom of the page, there is a pagination control showing '显示1到10,共196记录' and '第1页,共20页'.

模型名称	模型状态	模型分类	创建方式	创建人	来源	最新版本	更新时间	上线时间	下线时间	操作
高辉测试	未上线	似的发射点	在线方式	管理员	bonc		2018-04-26 10:03:00			上线 下线 建模 修改 删除 版本
ceshiyong	已下线	我的测试1	在线方式	管理员	bonc	v2	2018-04-25 15:37:13	2018-04-25 14:32:21	2018-04-25 14:32:24	上线 下线 建模 修改 删除 版本
调度测试流程-liuhao	已上线	我的测试1	在线方式	刘果	基础研发部	v1	2018-04-25 11:06:53	2018-04-25 11:06:54		上线 下线 建模 修改 删除 版本
最小哥斯基	未上线	我的测试1	在线方式	管理员	bonc		2018-04-23 15:42:59			上线 下线 建模 修改 删除 版本
新增2.0测试模型	未上线	我的测试1	离线方式	陈国	bonc		2018-04-23 14:07:27			上线 下线 建模 修改 删除 版本
v2测试	未上线	我的测试1	在线方式	管理员	bonc		2018-04-23 09:56:40			上线 下线 建模 修改 删除 版本
微微	未上线	我的测试1	在线方式	管理员	bonc		2018-04-23 09:25:40			上线 下线 建模 修改 删除 版本
new_week	未上线	我的测试1	在线方式	管理员	bonc		2018-04-23 08:53:16			上线 下线 建模 修改 删除 版本
离线	未上线	似的发射点	离线方式	管理员	bonc		2018-04-20 15:15:04			上线 下线 建模 修改 删除 版本
qwerdf	未上线	我的测试1	离线方式	管理员	bonc		2018-04-20 15:06:01			上线 下线 建模 修改 删除 版本

3.3.1.3 操作步骤

1、模型搜索

支持按模型名称、模型状态、模型分类、创建方式、更新日期的单一条件或组合条件的搜索，输入搜索条件后，点击右侧的“搜索”按钮即可。

2、 模型列表操作

支持对列表按照模型名称及更新时间的排序。

在列表中每一个模型的右侧，根据权限设置可对模型执行上线、下线、建模、修改、删除、版本的操作。

3.3.1.4 注意事项

上线后的模型，不能进行建模、修改、删除的操作。如需修改和建模，请先进行下线

3.3.2模型创建

3.3.2.1 业务描述

用户可通过此功能进行模型创建，对模型基本信息、说明信息、应用信息进行维护，并提供两种建模的入口，在线方式和离线方式。

3.3.2.2 功能界面

The screenshot displays the 'Model Creation' interface within the BONC Model Data Management System. The page title is '模型数据管理系统' and the user is logged in as '欢迎您, 管理员'. The main content area is titled '基本信息' and contains the following form fields:

- *模型名称: 新增模型
- *模型分类: 模型类别
- *业务类型: C网
- 模型描述: 不能超过500个字符

At the bottom of the form, there are two buttons: '下一步' (Next Step) and '返回' (Return).

3.3.2.3 操作步骤

从模型列表的右上方点击“创建模型”，进入到模型创建的页面。依次填写基本信息、

模型名称、模型分类、业务类型、模型描述等信息的内容。

在模型应用信息中，选择共享，则在上线后，可由管理员进行授权，被其他用户所见及使用；选择不共享，则在上线后，不能进行授权操作。

模型信息填写完成后，点击保存按钮，模型被保存，可在模型列表中可见并维护。

3.3.2.4 注意事项

模型名称在一个企业账号内不允许重复。

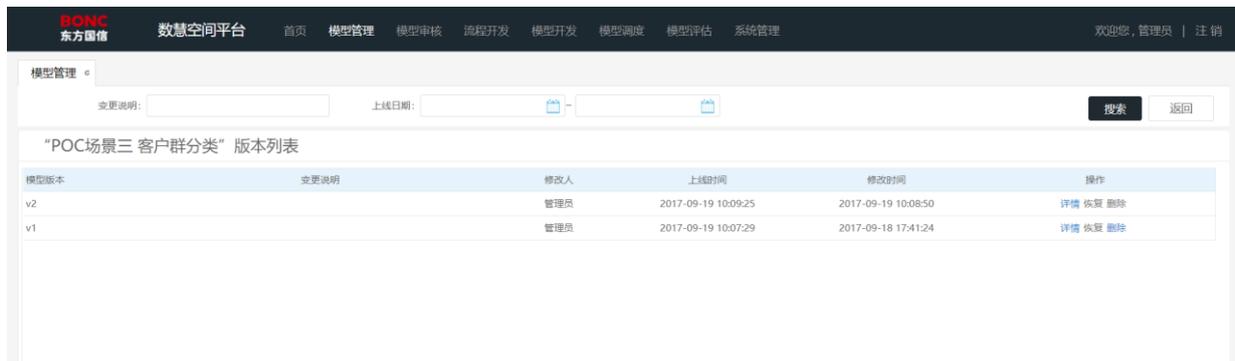
上传的文档，不能超过要求的限制大小。

3.3.3 模型版本管理

3.3.3.1 业务描述

模型在上线申请通过后，会生成新的版本，系统提供对历史版本的保存，用户可查看模型的演变过程。

3.3.3.2 功能界面



3.3.3.3 操作步骤

在模型列表中每一个模型的右侧，点击“版本”按钮，可进入到模型版本页面，查看模型的历史版本列表。

可在上面根据变更说明及上线日期对历史版本进行搜索。

可打开模型的任一版本，查看详情。

可对历史版本进行恢复及删除。

3.3.3.4 注意事项

模型的每一次上线申请通过后，都会生成新的版本，包括第一次创建后上线及后续的修改上线。未上线的模型不会形成版本。

3.3.4 模型开发

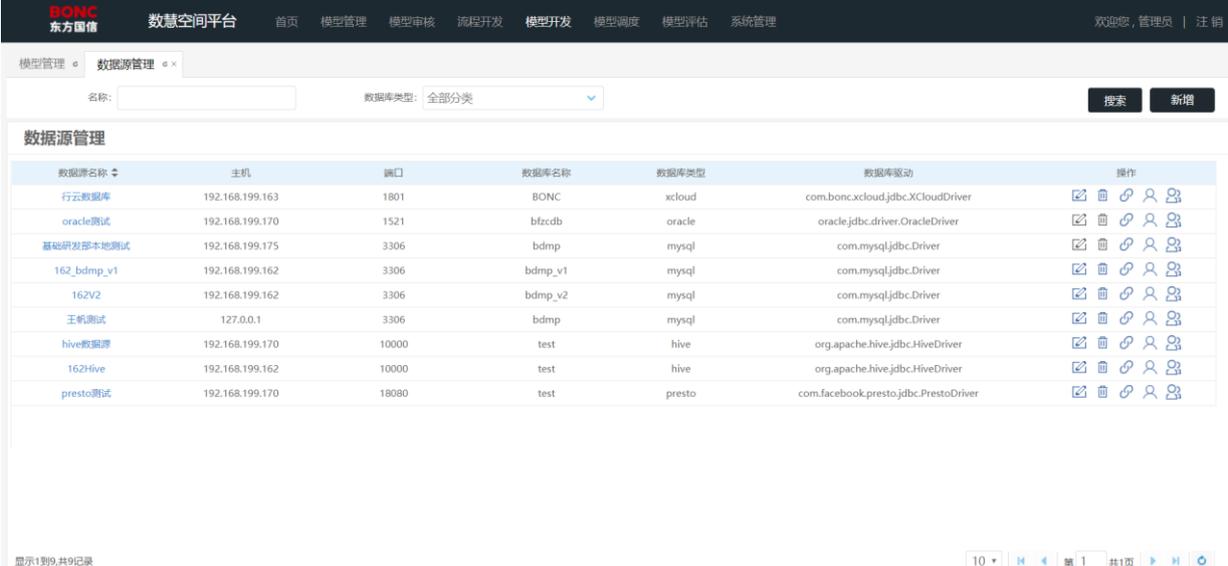
3.3.4.1 数据源管理

3.3.4.1.1 业务描述

对应菜单栏——模型开发-数据源管理。

配置数据库数据源，支持 Oracle、MySQL、Hive、Presto、行云类型的数据源配置。数据源配置用于加载模型挖掘的数据，及指定结果生成的路径。

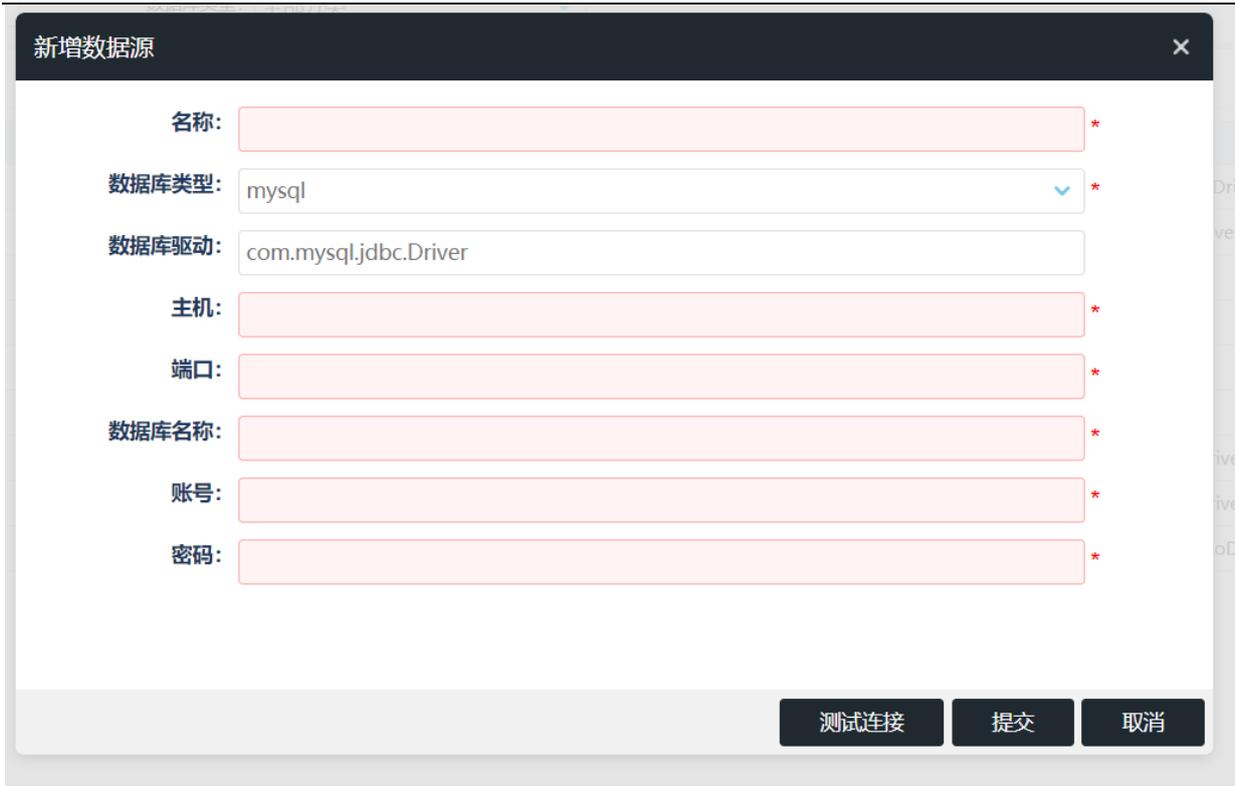
3.3.4.1.2 功能界面



The screenshot displays the '数据源管理' (Data Source Management) interface. At the top, there is a search bar and a dropdown menu for '数据库类型' (Database Type) set to '全部分类'. Below the search bar is a table with the following columns: '数据源名称' (Data Source Name), '主机' (Host), '端口' (Port), '数据库名称' (Database Name), '数据库类型' (Database Type), '数据库驱动' (Database Driver), and '操作' (Actions). The table contains 10 rows of data sources.

数据源名称	主机	端口	数据库名称	数据库类型	数据库驱动	操作
行云数据库	192.168.199.163	1801	BONC	xcloud	com.bonc.xcloud.jdbc.XCloudDriver	[Icons]
oracle测试	192.168.199.170	1521	bfzcdb	oracle	oracle.jdbc.driver.OracleDriver	[Icons]
基础研发部本地测试	192.168.199.175	3306	bdmp	mysql	com.mysql.jdbc.Driver	[Icons]
162_bdmp_v1	192.168.199.162	3306	bdmp_v1	mysql	com.mysql.jdbc.Driver	[Icons]
162V2	192.168.199.162	3306	bdmp_v2	mysql	com.mysql.jdbc.Driver	[Icons]
王帆测试	127.0.0.1	3306	bdmp	mysql	com.mysql.jdbc.Driver	[Icons]
hive数据源	192.168.199.170	10000	test	hive	org.apache.hive.jdbc.HiveDriver	[Icons]
162Hive	192.168.199.162	10000	test	hive	org.apache.hive.jdbc.HiveDriver	[Icons]
presto测试	192.168.199.170	18080	test	presto	com.facebook.presto.jdbc.PrestoDriver	[Icons]

At the bottom of the table, there is a pagination control showing '显示1到9,共9记录' (Display 1 to 9, total 9 records) and a page number '1' out of '1'.



新增数据源

名称: *

数据库类型: mysql *

数据库驱动: com.mysql.jdbc.Driver

主机: *

端口: *

数据库名称: *

账号: *

密码: *

测试连接 提交 取消

3.3.4.1.3 操作步骤

进入数据源管理，点击“新增数据源”，配置数据库类型、驱动、主机、端口、账号、密码信息。

配置完成后，可测试连接，通过后，点击提交保存该数据源配置。

配置完成的数据源，可进行编辑、删除、测试连接、用户授权、企业账号授权。被授权的用户才能从系统中连接该数据源。

3.3.4.1.4 注意事项

用户在配置数据库数据集前，需先保障数据源管理中已连接所属数据源。

用户在指定挖掘结果生成的数据表前，需先保障数据源管理中已连接所属数据源。

单机服务器版，不支持 Hive 的数据源。分布式云版，支持 Hive 的数据源。

3.3.4.2 数据集管理

3.3.4.2.1 业务描述

对应菜单栏——模型开发-数据集。

支持磁盘文件的数据集、HDFS 的数据集，以及通过 SQL 输入的方式，将所需数据加载至挖掘建模工具在模型训练及模型执行时所用到的数据集中。

3.3.4.2.2 功能界面

数据集名称	数据集类型	数据源名称	SQL语句	文件路径	创建人	创建时间	修改人	修改时间	操作
ceshi	磁盘文件			datasetFile/1524708239583/timeseries.xls	管理员	2018-04-26 10:03:58	管理员	2018-04-26 10:03:58	编辑 删除 授权
数据源	HDFS文件			datasetFile/1524473740987/timeseries.xls	管理员	2018-04-23 04:55:40	管理员	2018-04-23 04:55:40	编辑 删除 授权
是否换机数据集	数据库	162V2	select * from data_change limit 9999		管理员	2018-04-23 04:32:37	管理员	2018-04-23 04:32:37	编辑 删除 授权
162房价	数据库	162V2	select * from t_boston_house_prices		管理员	2018-04-18 09:29:32	管理员	2018-04-18 09:29:32	编辑 删除 授权
推荐算法数据集	数据库	162V2	select * from t_shopping_item		管理员	2018-04-17 10:51:21	管理员	2018-04-17 10:51:21	编辑 删除 授权
oracle	数据库	oracle测试	select * from test_t_6434		管理员	2018-03-29 09:03:58	管理员	2018-04-16 05:10:15	编辑 删除 授权
hive_from_mysql_label	数据库	162Hive	select * from test_t_7123		管理员	2018-04-16 03:22:50	管理员	2018-04-16 03:22:50	编辑 删除 授权
mysql小写label	数据库	162V2	select * from classify_label_low		管理员	2018-04-16 03:17:24	管理员	2018-04-16 03:17:24	编辑 删除 授权
ORACLE_unlabel	数据库	oracle测试	select * from CLASSIFY_UNLABELED		管理员	2018-04-16 11:04:21	管理员	2018-04-16 11:04:21	编辑 删除 授权
ORACLE_label	数据库	oracle测试	select * from CLASSIFY_LABEL		管理员	2018-04-16 11:02:45	管理员	2018-04-16 11:02:45	编辑 删除 授权

新增数据集

数据集名称:

数据集类型:

数据源:

sql:

3.3.4.2.3 操作步骤

进入数据集管理，点击“新增数据集”，填写数据集名称，选择数据集类型；如果为磁盘文件数据集或 HDFS 文件数据集，则需要选择本地文件进行上传；如果为数据库数据集，则需要选择在数据源中已配置好的数据库，再撰写 SQL 指定表数据。完成后，点击提交保存数据集。保存后，即可以在可视化建模时，使用该数据集进行模型训练，也可以在新建调度时选择该数据集进行模型应用。

配置完成的数据集，可进行编辑、删除、用户授权、企业账号授权。被授权的用户才能从系统中使用该数据集。

3.3.4.2.4 注意事项

分布式云版，不支持磁盘文件的数据源。

3.3.4.3 离线模型日志管理

3.3.4.3.1 业务描述

对应菜单栏——模型开发-离线模型日志管理。

离线模型日志配置可进行新增，修改和删除。

3.3.4.3.2 功能界面

模型管理 离线模型日志管理

日志名称: 调度工具类型: 全部分类

离线模型日志配置列表

日志配置名称	调度工具类型	修改时间	修改人	日志数据源	日志信息表	存储过程ID/作业ID	存储过程描述/作业名称	周期列	执行状态列	执行开始时间列	执行结束时间列	操作
无物理表测试	Oracle	2018-01-11 10:45:45	管理员	presto测试	test	111	2222	3333	444	555555	6666666666666666	修改 删除
162MySQL-log	ETL			162V2	add_info	job_id	job_name	acct_month	code_desc	time1	time2	修改 删除

日志配置名称: 调度工具类型: 请选择
 日志数据源: 请选择数据源 日志信息表:

ID对应列: 名称对应列: 周期对应列:
 执行状态对应列: 执行开始时间对应列: 执行结束时间对应列: 该输入项为必填项

3.3.4.3.3 操作步骤

进入离线模型日志配置，点击“新增”，即可填写离线模型日志配置信息

配置完成的离线模型日志，可进行修改、删除。

3.3.4.3.4 注意事项

3.3.4.4 模型开发-离线方式

3.3.4.4.1 业务描述

模型的两种建模方式之一，适用于已在线下完成模型的开发脚本，需通过数慧空间提

供的线上管理手段，沉淀模型的知识库。

3.3.4.4.2 功能界面

模型管理

离线模型脚本信息

脚本名称:

脚本类型: 过程脚本 Python脚本 R脚本

建模算法:

建模工具: sql python R

运行环境:

关键字:

建议销售品:

建议销售渠道:

选择脚本文件:

模型管理

模型输出信息

输出标识:

数据源:

表名:

字段名	字段中文名	字段类型	标记
return data type	return data type	VARCHAR	地市
api code	api code	VARCHAR	用户ID
sql	sql	TEXT	账期
api name	api name	VARCHAR	用户手机号

模型管理

模型执行信息

离线模型日志配置选择:

执行周期:

日志数据源: 基础研发部本地测试

日志表: classify_label_new

日志表结构:

存储过程ID: ACCT_MONTH	存储过程名称: SERV_ID
参数: AVG_DURATION_5	执行状态: DURATION_5_TRANS
执行开始时间: PRESENT_EVDO_FLOW	执行结束时间: PRESENT_EVDO_FLOW

3.3.4.4.3 操作步骤

在模型列表中，选择要建模的模型，点击模型后方的“建模”按钮，在弹出窗选择“离线开发”。

在本地选择已完成的建模脚本，填写脚本名称，运行环境等内容，之后选择该脚本的输出信息以及日志信息，点击提交，进行脚本上传及保存。

3.3.4.4.4 注意事项

离线脚本文件大小不能超过要求的限制大小。

日志信息需要在 离线模型日志管理中配置。

输出信息默认必须配置用户 id，手机号，账期和地市这 4 个信息。

3.3.4.5 模型开发-在线

3.3.4.5.1 业务描述

模型的两种建模方式之一，支持不同行业、不同场景、不同业务目标的数据挖掘需求，通过算法组件拖拽、作业引导、可视化数据探索的方式，使业务人员轻松、简便的创建模型。

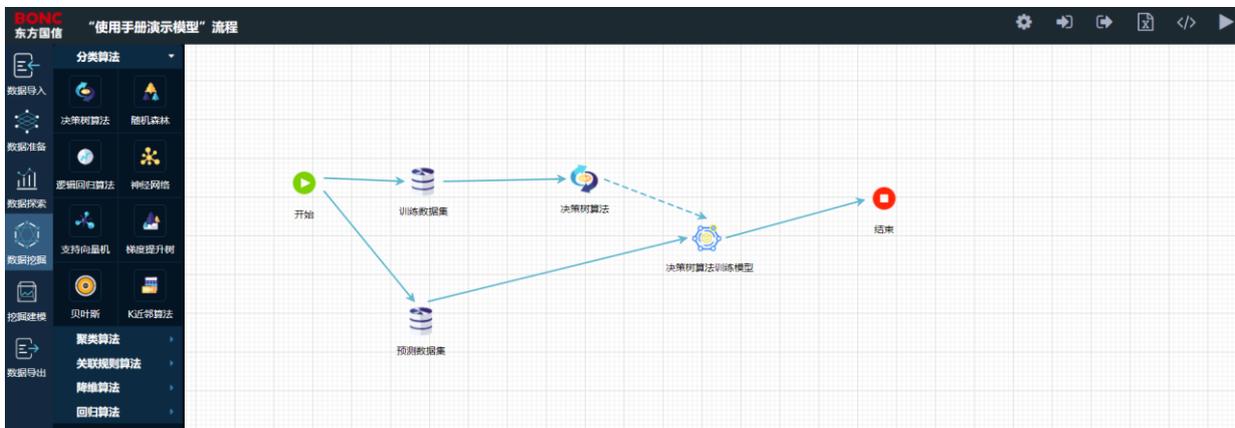
下图为目前所支持的算法节点：







3.3.4.5.2 功能界面



3.3.4.5.3 操作步骤

在模型列表中，选择要建模的模型，点击模型后方的“建模”按钮，在弹出窗选择“在线开发”。

- 节点拖拽

进入可视化挖掘建模工具中，左侧为挖掘建模工具库，右侧为建模画布，可根据建模需要，将左侧任一算法节点拖拽至画布中。

- 参数配置

拖拽至画布中的节点，可双击或右键点击编辑打开参数配置页面，各个算法组件支持详细的参数调整。

节点定义	特征列	标签列
节点名称	决策树算法	
训练数据集	训练数据集数据	
测试集占比	0.3	
计算方法	<input checked="" type="radio"/> 基尼系数 <input type="radio"/> 信息增益	
决策树最大深度	5	
节点备注		

- 数据挖掘

针对数据挖掘等复杂算法，在节点上右键点击“训练模型”，可以对模型进行训练，产生相应算法的训练模型，在将预测数据与训练模型相连，即可进行预测。



对训练的模型满意，可将模型保存作为预测模型在预测模型节点中使用。



- 预测模型

预测模型算法节点，可以用已有的分享的训练模型进行数据预测，并可查看预测模型的效果和参数配置等情况

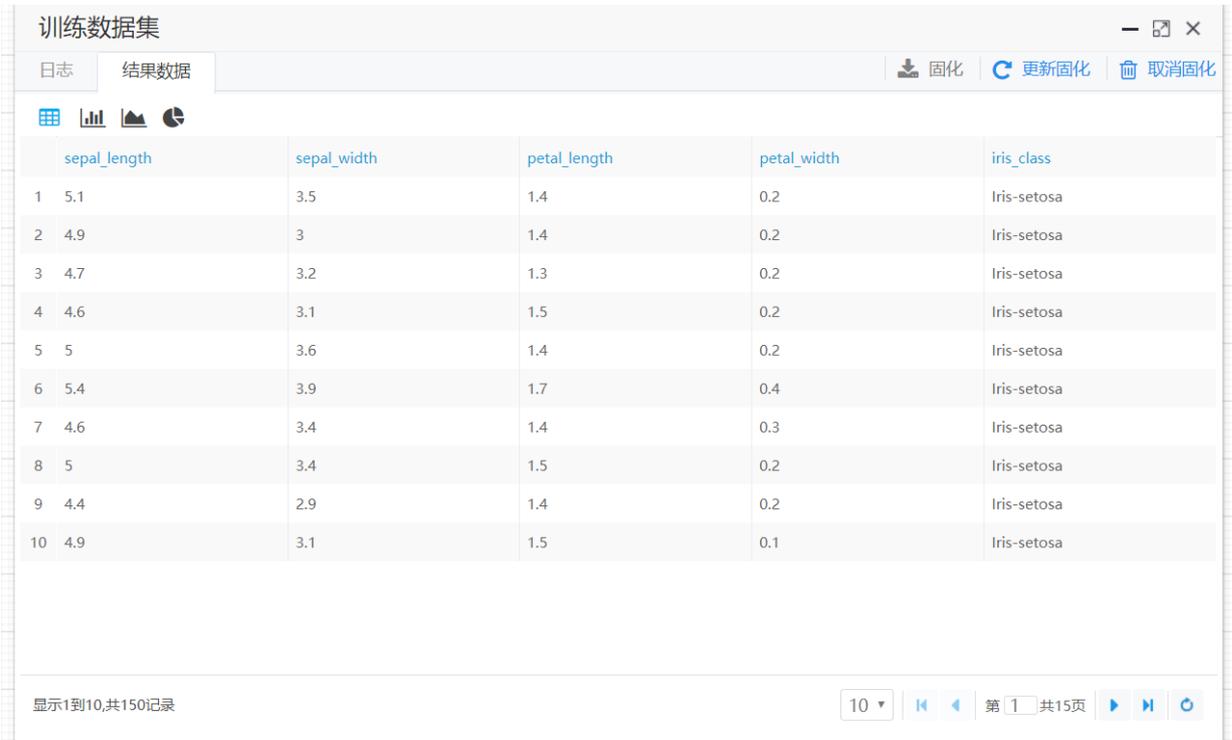


- 可视化探索分析

针对挖掘建模的完整流程，在节点上右键点击“预览数据”按钮，可以查看任一节点

的结果，并可以进行可视化的交互式探索，帮助用户更生动的分析错综复杂的数据，以便更有效的构建模型。

支持图标、直方图、线图、饼图的展示方式，并可变换维度和指标进行交互式的探索分析。

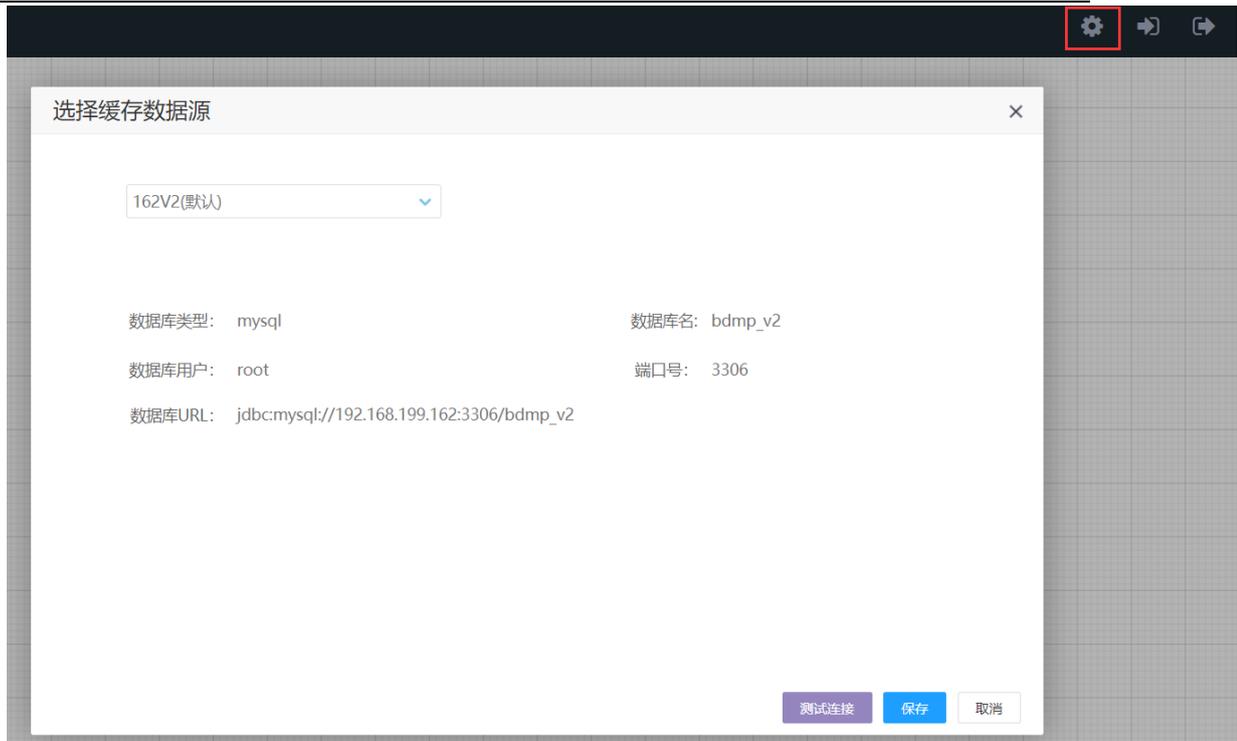


	sepal_length	sepal_width	petal_length	petal_width	iris_class
1	5.1	3.5	1.4	0.2	Iris-setosa
2	4.9	3	1.4	0.2	Iris-setosa
3	4.7	3.2	1.3	0.2	Iris-setosa
4	4.6	3.1	1.5	0.2	Iris-setosa
5	5	3.6	1.4	0.2	Iris-setosa
6	5.4	3.9	1.7	0.4	Iris-setosa
7	4.6	3.4	1.4	0.3	Iris-setosa
8	5	3.4	1.5	0.2	Iris-setosa
9	4.4	2.9	1.4	0.2	Iris-setosa
10	4.9	3.1	1.5	0.1	Iris-setosa

● 功能介绍

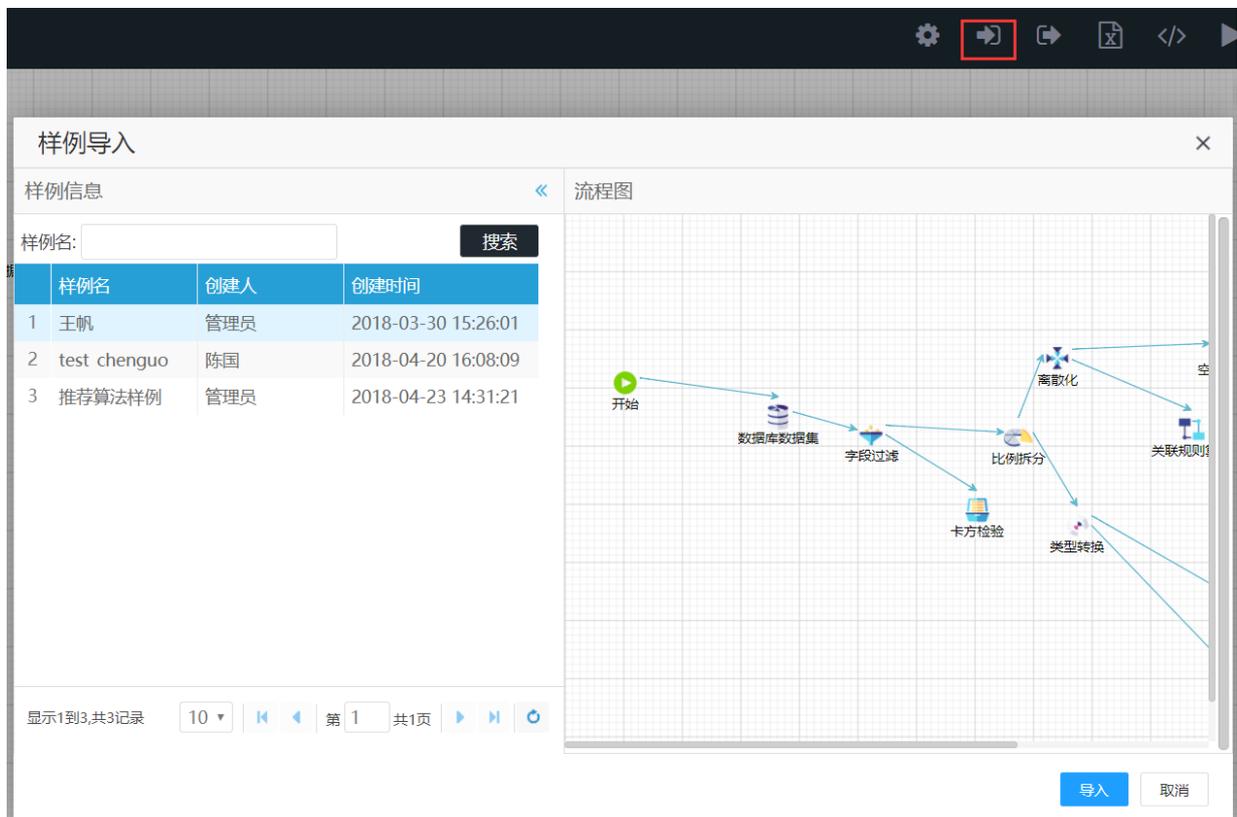
1. 缓存数据源选择

点击页面右上方工具栏的齿轮图标（第一个图标），会弹出缓存数据源的选择，可根据已经配置的数据源去选择要将此流程所缓存和固化的表存放到一个数据库。



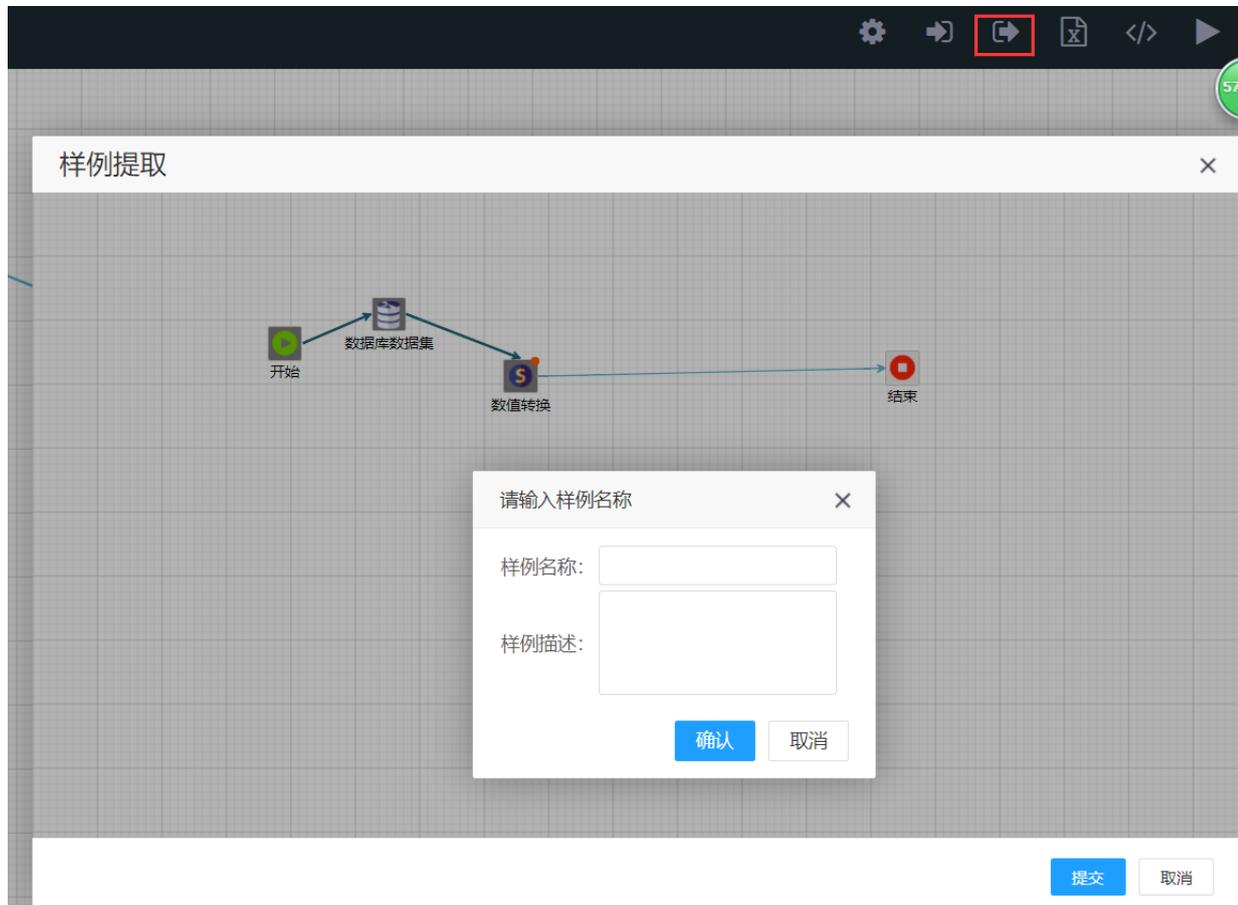
2. 样例库导入

点击页面右上方工具栏的导入图标（第二个图标），会弹出样例流程的选择，可根据已经存在的样例流程去选择要在此流程中导入哪一个样例流程。



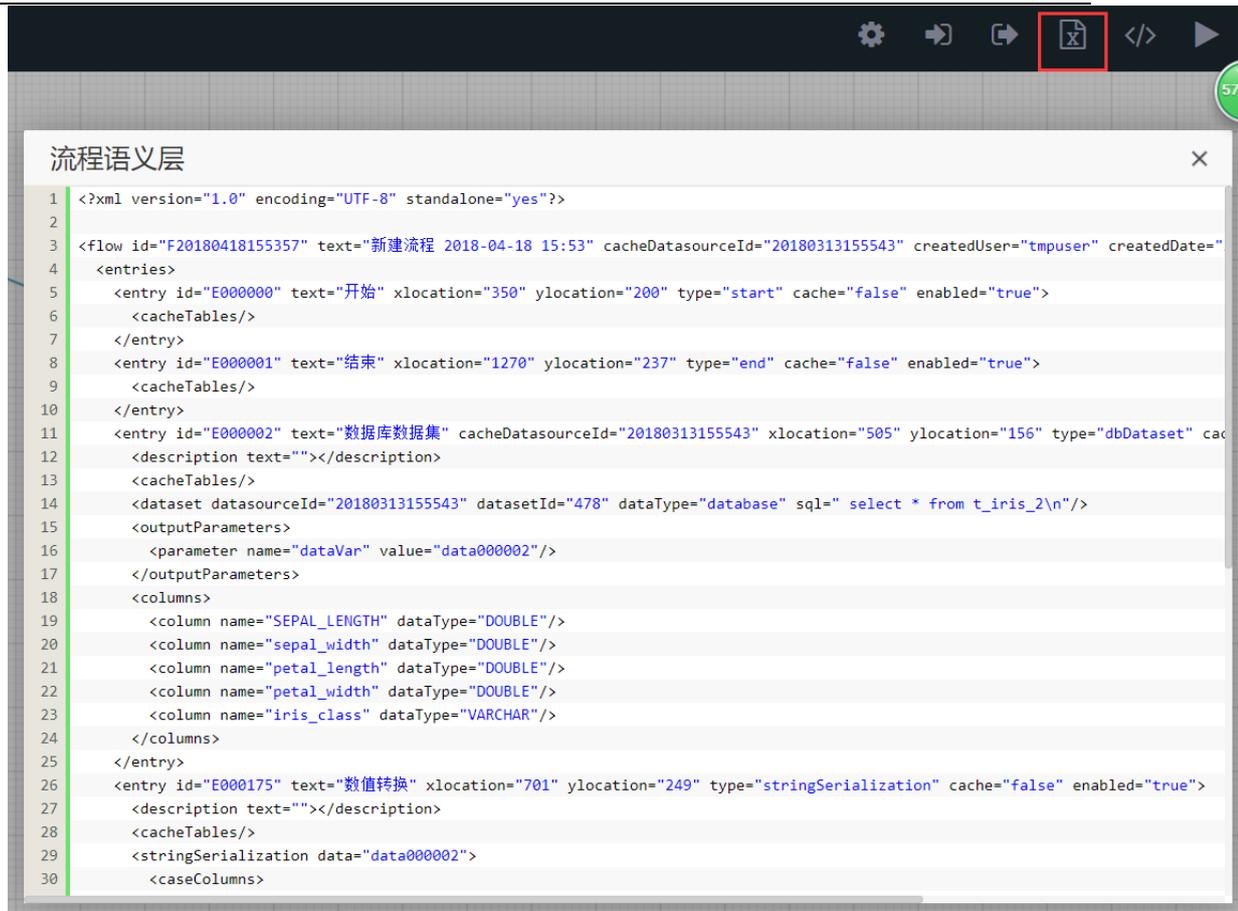
3. 样例库导出

点击页面右上方工具栏的导入图标（第三个图标），会弹出样例提取的选择，选择想要提取的流程，依次点击节点,变暗的节点就可提取至样例库,之后输入样例名称即可



4. 查看语义层

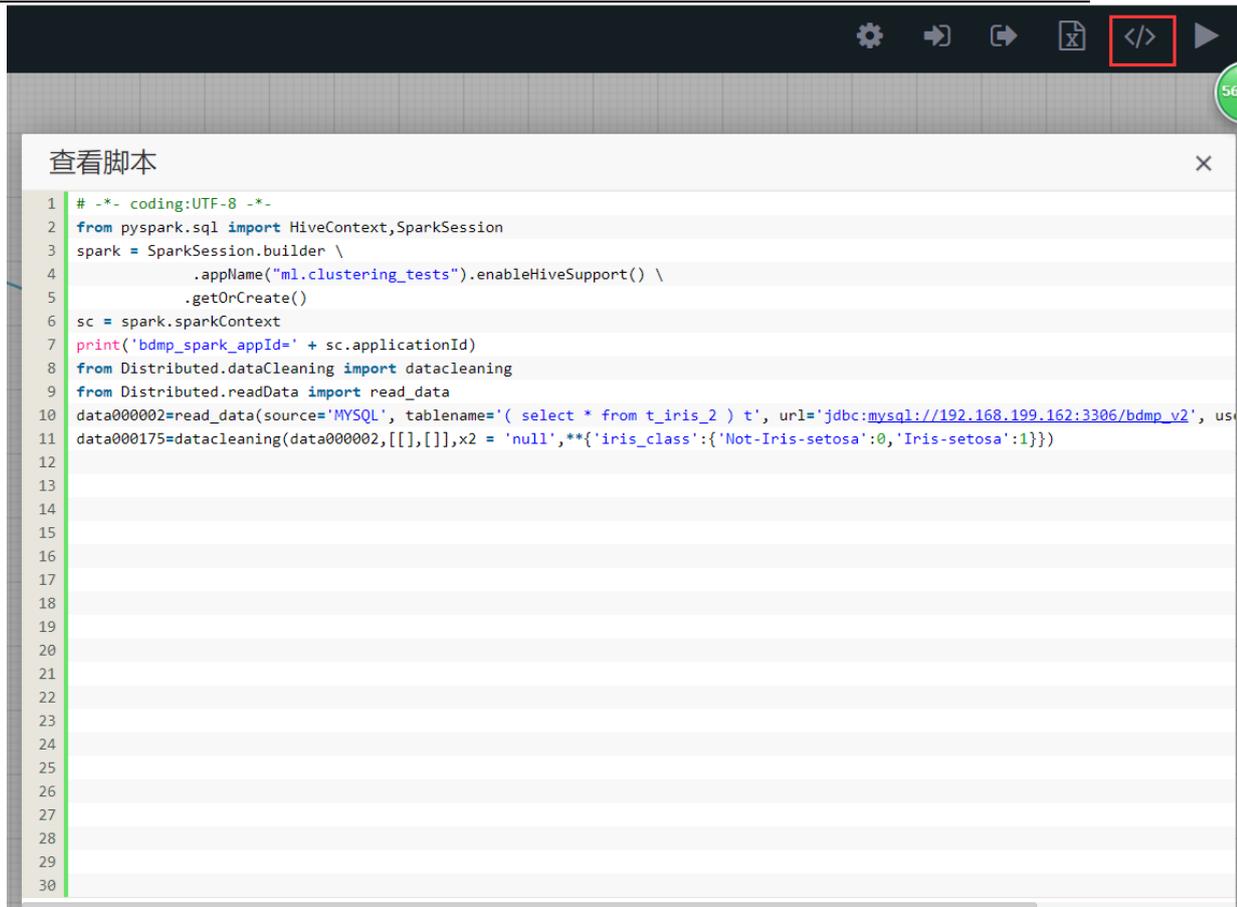
点击页面右上方工具栏的导入图标（第四个图标），会弹出语义层的信息，在此即可查看语义层结构



```
1 <?xml version="1.0" encoding="UTF-8" standalone="yes"?>
2
3 <flow id="F20180418155357" text="新建流程 2018-04-18 15:53" cacheDatasourceId="20180313155543" createdUser="tmpuser" createdDate="
4 <entries>
5   <entry id="E000000" text="开始" xlocation="350" ylocation="200" type="start" cache="false" enabled="true">
6     <cacheTables/>
7   </entry>
8   <entry id="E000001" text="结束" xlocation="1270" ylocation="237" type="end" cache="false" enabled="true">
9     <cacheTables/>
10  </entry>
11  <entry id="E000002" text="数据库数据集" cacheDatasourceId="20180313155543" xlocation="505" ylocation="156" type="dbDataset" cac
12    <description text=""></description>
13    <cacheTables/>
14    <dataset datasourceId="20180313155543" datasetId="478" dataType="database" sql=" select * from t_iris_2\n"/>
15    <outputParameters>
16      <parameter name="dataVar" value="data000002"/>
17    </outputParameters>
18    <columns>
19      <column name="SEPAL_LENGTH" dataType="DOUBLE"/>
20      <column name="sepal_width" dataType="DOUBLE"/>
21      <column name="petal_length" dataType="DOUBLE"/>
22      <column name="petal_width" dataType="DOUBLE"/>
23      <column name="iris_class" dataType="VARCHAR"/>
24    </columns>
25  </entry>
26  <entry id="E000175" text="数值转换" xlocation="701" ylocation="249" type="stringSerialization" cache="false" enabled="true">
27    <description text=""></description>
28    <cacheTables/>
29    <stringSerialization data="data000002">
30      <caseColumns>
```

5. 查看脚本

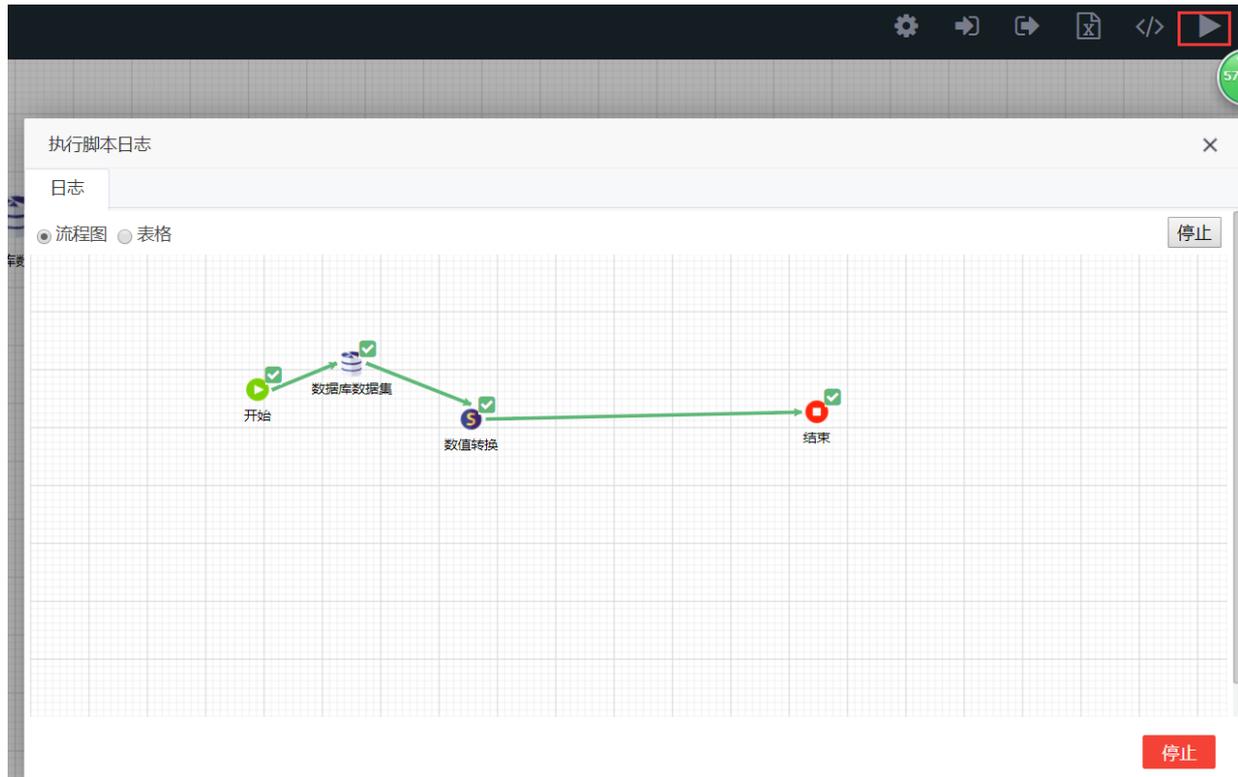
点击页面右上方工具栏的导入图标（第五个图标），会弹出生成脚本的信息，在此即可查看被执行的脚本



```
1 # -*- coding:UTF-8 -*-
2 from pyspark.sql import HiveContext,SparkSession
3 spark = SparkSession.builder \
4     .appName("ml.clustering_tests").enableHiveSupport() \
5     .getOrCreate()
6 sc = spark.sparkContext
7 print('bdmp_spark_appId' + sc.applicationId)
8 from Distributed.dataCleaning import datacleaning
9 from Distributed.readData import read_data
10 data000002=read_data(source='MYSQL', tablename='( select * from t_iris_2 ) t', url='jdbc:mysql://192.168.199.162:3306/bdmp_v2', us
11 data000175=datacleaning(data000002,[],[],x2 = 'null',**{'iris_class':{'Not-Iris-setosa':0,'Iris-setosa':1}})
12
13
14
15
16
17
18
19
20
21
22
23
24
25
26
27
28
29
30
```

6. 运行

点击页面右上方工具栏的导入图标（第六个图标），会执行连接到结束节点的所有流程



- 结果导出

结果的输出包括挖掘结果和评估结果，支持文件导出和数据库导出。

文件导出需对文件名称进行命名，并选择是否加密。

数据库导出需先配置数据源，再选择数据表。

3.3.4.5.4 注意事项

部分算法的参数提供默认的参数值，也可以进行自定义参数值。

可选择执行完整流程，也可选择执行到任一节点。

执行时系统展示执行日志，用户可查看发现是否有执行错误。

想要查看脚本，需将节点连至结束节点

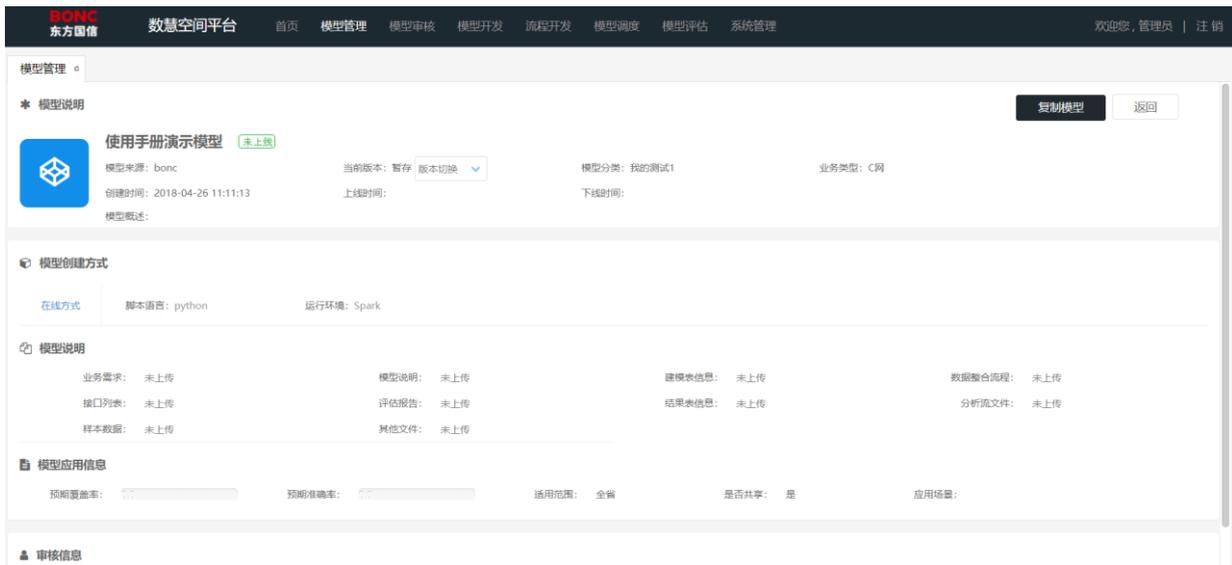
预测模型节点的查看算法和查看效果，需要选择预测模型保存后生效

3.3.5 模型详情

3.3.5.1 业务描述

在模型管理的列表及模型审核的列表中，可以点击模型名称，查看模型详情。

3.3.5.2 功能界面



3.3.5.3 操作步骤

入口 1: 模型管理-点击模型名称, 进入模型详情。

入口 2: 模型审核-审核-点击模型名称, 进入模型详情。

可查看模型详情、应用信息, 下载模型说明文档。

离线方式创建的模型, 可下载离线脚本包;

在线方式创建的模型, 可查看可视化在线流程。

由于一个模型会因为修改上线, 生成多个模型版本, 可在上方切换模型版本, 进行查看。

右上方的“复制模型”按钮, 用于创建一份同样的包含在线/离线建模信息的模型, 方便用户基于原有模型进行修改后再生成一个新的模型。

3.3.5.4 注意事项

版本切换时, 未上线版本会被命名为“暂存”, 也可以进行切换查看。

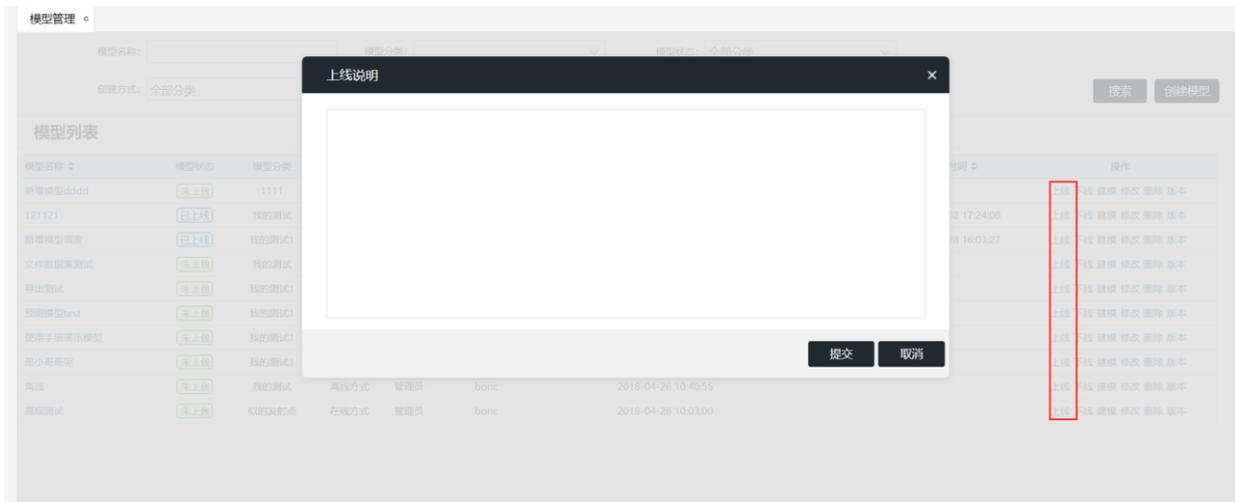
复制模型时, 会复制模型的基本信息、应用信息、可视化在线挖掘流程, 不会复制附件信息。

3.3.6 模型上下线

3.3.6.1 业务描述

模型创建完成后，可进行模型上线申请，上线后的模型会生成新的版本，并被其他用户所见及调用。在需要对模型进行修改，或者没有在线价值时，可对模型执行下线操作。

3.3.6.2 功能界面



3.3.6.3 操作步骤

- 上线申请

在模型列表中，选择要上线的模型，点击模型后方的“上线”按钮，在弹出窗填写上线说明，将上线申请提交至管理员。

- 模型上线

模型的上线申请，被管理员通过后，模型的状态从“已创建”变更为“已上线”，并生成新的版本。处于上线状态的模型不能进行修改、删除、建模的操作。

- 模型下线

在模型列表中，选择要下线的模型，点击模型后方的“下线”按钮，模型即会下线，下线无需管理员审批。

3.3.6.4 注意事项

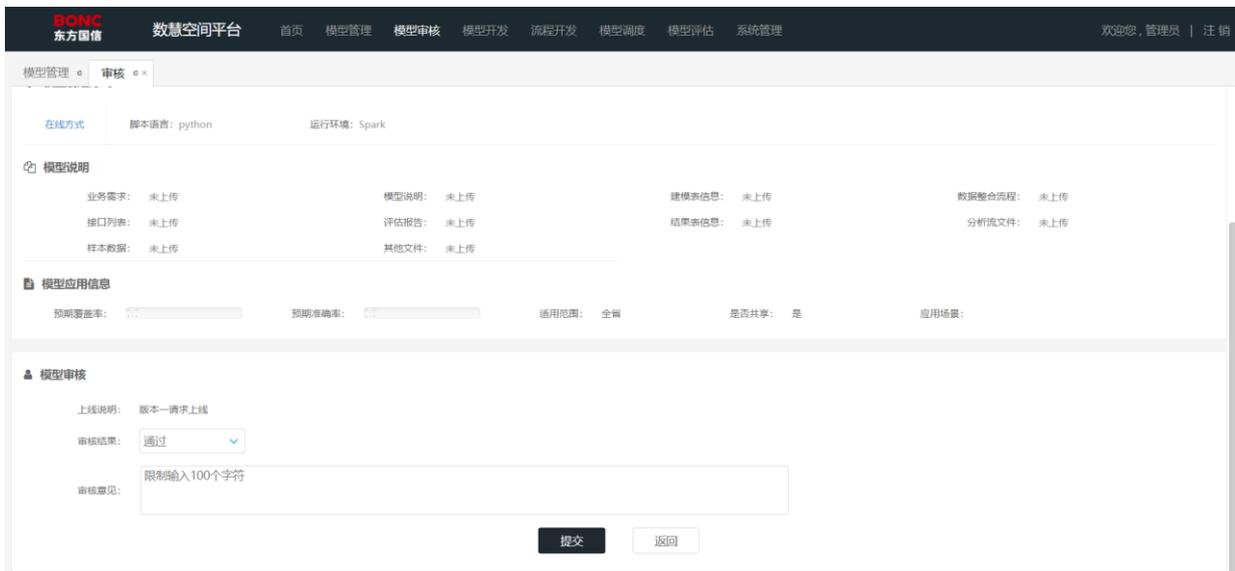
被调用的模型，不会受下线的影响，也不会受版本变更的影响。

3.3.7 模型审核及授权

3.3.7.1 业务描述

具备审核权限的用户，会收到模型的上线申请，决定是否将该模型上线，并在上线后，对模型进行授权。

3.3.7.2 功能界面



3.3.7.3 操作步骤

- 模型审核

进入菜单栏“模型审核-审核”，看到未审核及已审核的模型列表，选择标注为“未审核”模型，点击“未审核”按钮，进入到模型审核页面。浏览模型信息、上线说明，可对附件进行下载，之后管理员决定是否通过模型的上线申请，并填写审核意见。

- 模型授权

进入菜单栏“模型审核-授权”，看到所有已上线模型，选择上线模型进行人员授

权，以决定模型可以被哪些用户可见及使用。

授权包括人员授权及企业账号授权，可在用户列表中选择相应用户，企业账号（即租户）授权后，该企业账号下的所用用户都会被授权。

3.3.7.4 注意事项

模型审核通过后，模型状态会变更为已上线。

审核通过但未授权的模型，只能被创建人和管理员可见。

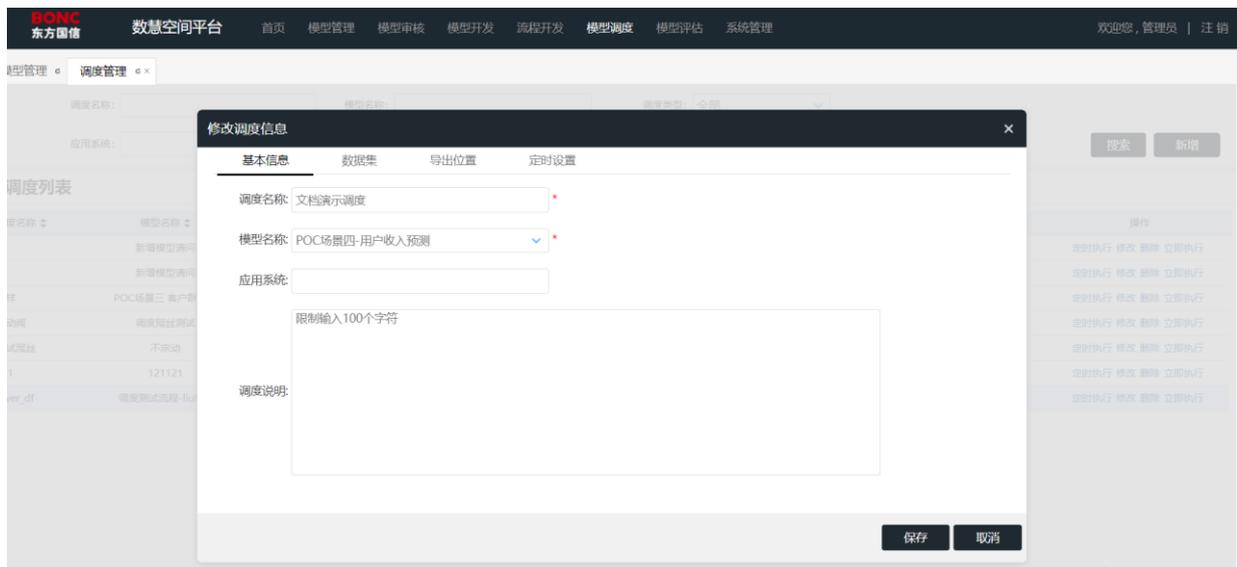
3.3.8 模型调度

3.3.8.1 业务描述

对应菜单栏——模型调度。

上线后的模型，可快速部署至生产系统中，可对接统一调度中心，实现对模型的直接引用，从模型训练到模型应用的平滑过渡。

3.3.8.2 功能界面



3.3.8.3 操作步骤

- 调度管理

进入菜单栏“模型调度-调度管理”页面，可点击“新建调度”，选择需要投入到

生产系统中的模型，填写调度名称。

系统会自动读取到原有模型训练时的数据集，可用户选择替换为生产系统中的数据集。

用户需指定挖掘结果生成的方式、路径。

完成调度时间、执行频率的设置。此步也可直接由统一调度中心设置。

点击“保存”，新建调度保存至调度管理的列表中。

用户在调度列表中，启动某一调度，即将模型快速投入到生产系统中

- 调度监控

进入菜单栏“模型调度-调度监控”页面，对已执行的调度进行监控。

可查看任务执行的状态，执行的开始时间及结束时间。

可对执行失败的任务，进行重新执行。

可查看任务的执行日志。

3.3.8.4 注意事项

模型被下线或有新的版本生成时，原有调度不受影响。

单机服务器版，无模型调度模块。分布式云版有此模块。

3.3.9 模型评估

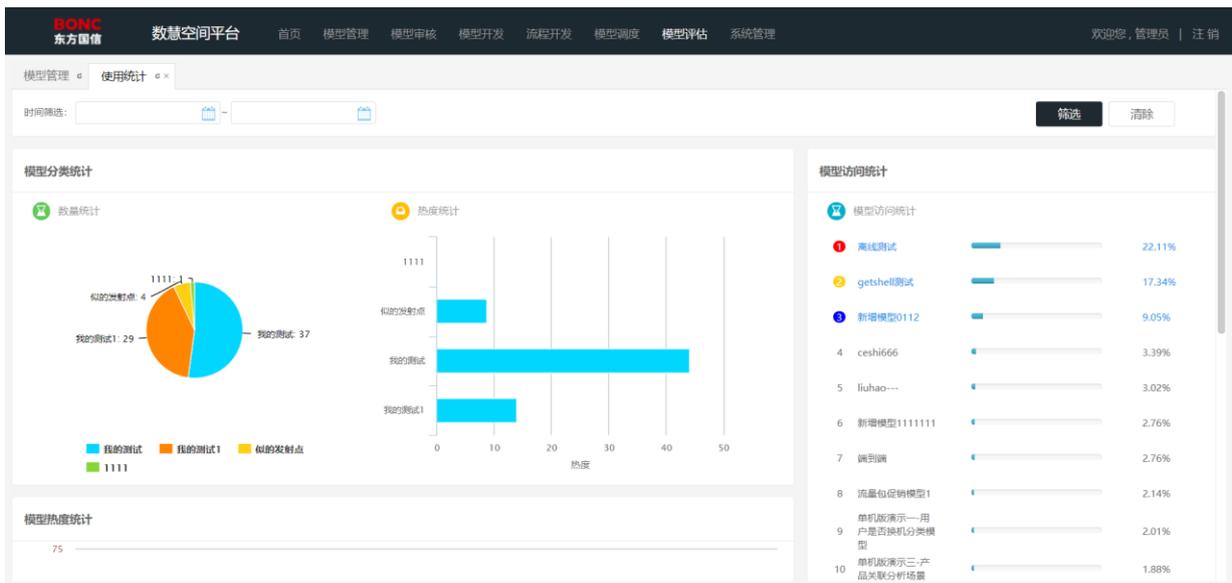
3.3.9.1 业务描述

对应菜单栏——模型评估。

模型使用统计：通过对模型热度分析，了解模型的使用价值。

模型效果评估：通过模型执行效果对模型进行优化。

3.3.9.2 功能界面



3.3.9.3 操作步骤

- 模型使用统计

进入菜单栏“模型评估-使用统计”页面，从模型分类、热度、访问、调用等维度，对模型的使用情况进行统计。

- 模型效果评估

进入菜单栏“模型评估-效果评估”页面，查看模型的执行情况，并通过回收的执行效果数据，对模型价值进行全面评估，以便决定是否需要对模型进行优化。

3.3.9.4 注意事项

模型效果评估的数据需要从外部系统采集。

3.4 系统管理

3.4.1 客户端监控

3.4.1.1 业务描述

对应菜单栏——系统管理-客户端监控。

数慧空间会同时启动多个数慧空间的服务，当某一台服务器宕机时，另外一个服务不会受到影响，这样保证了服务的稳定性，更提高了服务的高可用性。

此功能用于监控数慧空间部署的客户端。

3.4.1.2 功能界面



3.4.1.3 操作步骤

进入客户端监控，数慧空间采用服务分布式+计算分布式的方式，当前可查看客户端运行情况，包括客户端地址、缓冲池、线程数、内存占用、磁盘空间情况等信息。

3.4.1.4 注意事项

只适用于管理员权限的用户，关机操作需谨慎。

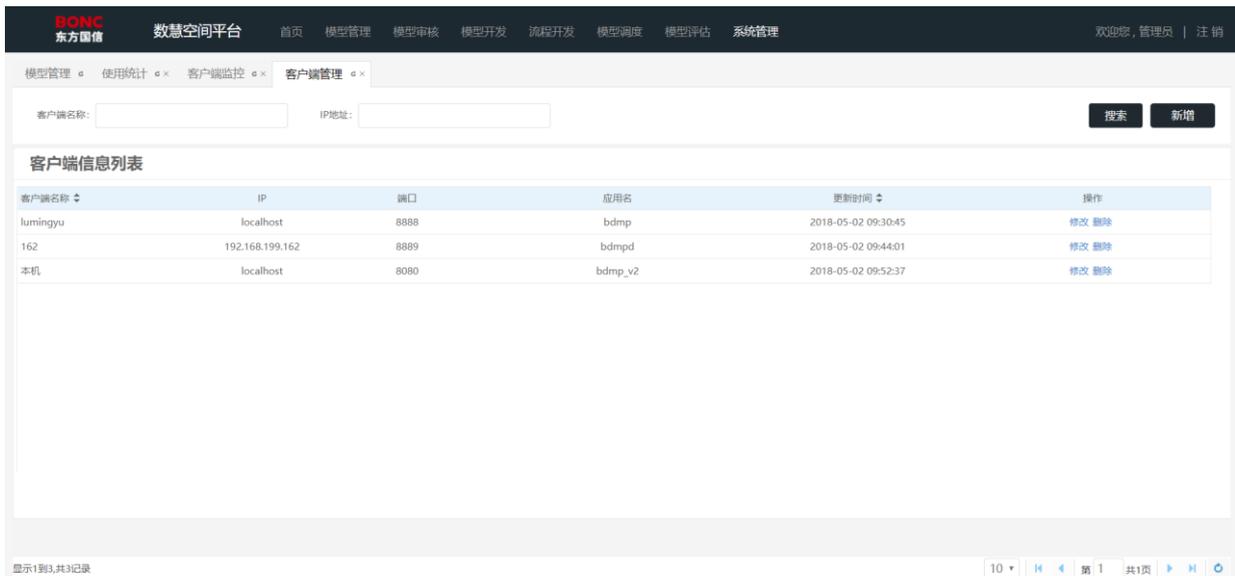
3.4.2 客户端管理

3.4.2.1 业务描述

对应菜单栏——系统管理-客户端管理。

适用于对客户端的管理。

3.4.2.2 功能界面



3.4.2.3 操作步骤

进入客户端管理，查看当前的客户端信息列表，并可以进行新增、修改、删除的操作。

客户端配置完成后，后台启动后，可进入到客户端监控中可以看到其运行状态。

3.4.2.4 注意事项

除需在前台进行客户端配置外，后台还需更新相应的配置文件。

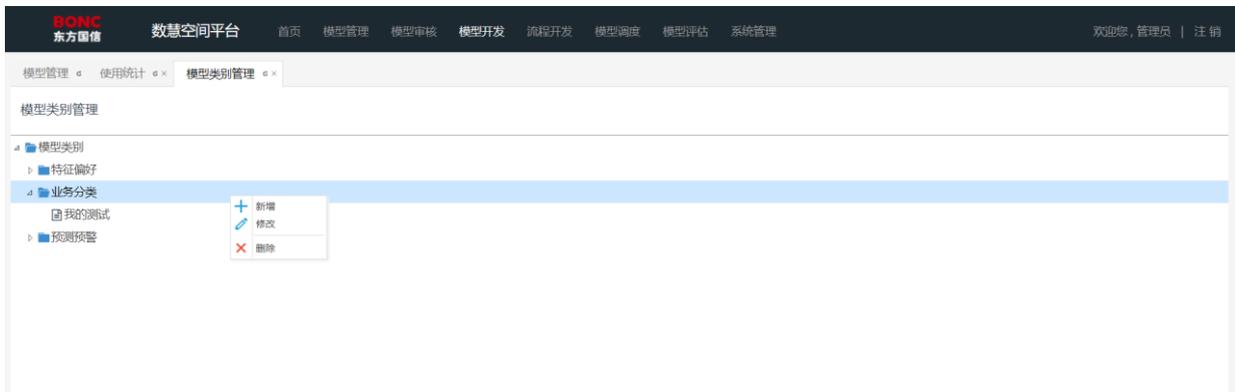
3.4.3 模型类别管理

3.4.3.1 业务描述

对应菜单栏——模型开发-模型类别管理。

对模型的类别进行管理维护，用于在模型创建时，模型类别的选择。

3.4.3.2 功能界面



3.4.3.3 操作步骤

采用树形结构管理“模型类别”。

可通过右键对模型类别进行新增、修改及删除。

维护信息包括分类名称、排序、分类编码，分类名称、排序需手工维护，分类编码自动生成。

3.4.3.4 注意事项

排序信息，指的是在模型创建时，模型类别选项中类别的展示顺序，如果此处将排序编号设为重复，系统会随机将排序号码重复的类别排序。

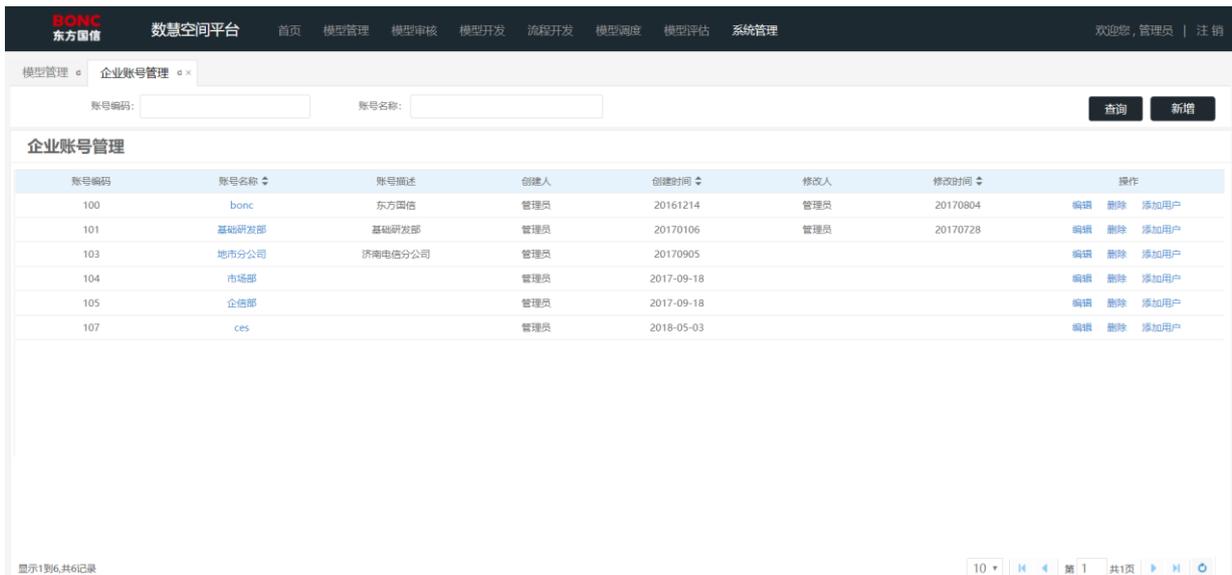
3.4.4 企业账号管理

3.4.4.1 业务描述

对应菜单栏——系统管理-企业账号管理。

企业账号，同租户的概念，企业账号间数据隔离。企业账号下可创建用户，用户可归属在多个企业账号下。

3.4.4.2 功能界面



3.4.4.3 操作步骤

进入企业账号管理，可查看企业账号列表，上方可进行搜索。

可新增企业账号，维护企业账号信息。

企业账号创建后，可进行添加用户的操作，将已创建的用户添加至该企业账号下。

3.4.4.4 注意事项

分配用户前，需先创建该用户。

用户可归属在多个企业账号下。

涉及到授权的操作，如果企业账号被授权，则该企业账号下的用户也会被授权。

3.4.5 用户管理

3.4.5.1 业务描述

对应菜单栏——系统管理-用户管理。

对用户信息及权限进行管理。

3.4.5.2 功能界面

The screenshot shows the '新增用户' (Add User) form in the BONC system. The form is titled '新增用户' and has a '保存' (Save) button and a '取消' (Cancel) button in the top right corner. The form fields are as follows:

登录 ID:	<input type="text"/>	用户姓名:	<input type="text"/>
登录密码:	<input type="password"/>	确认密码:	<input type="password"/>
管理员:	<input type="checkbox"/>	用户性别:	<input type="text" value="男"/>
移动电话:	<input type="text"/>	固定电话:	<input type="text"/>
邮箱地址:	<input type="text"/>	企业账户:	<input type="text" value="基础研发部,bonc"/>
账号状态:	<input type="text" value="启用"/>		
备注信息:	<input type="text"/>		

3.4.5.3 操作步骤

进入到用户管理，可查看系统中的用户列表。

可对新建用户，维护用户的 ID、密码、姓名、邮箱等信息。

用户创建后，可对用户的菜单权限进行设置。

3.4.5.4 注意事项

3.5 流程开发

3.5.1 预测模型管理

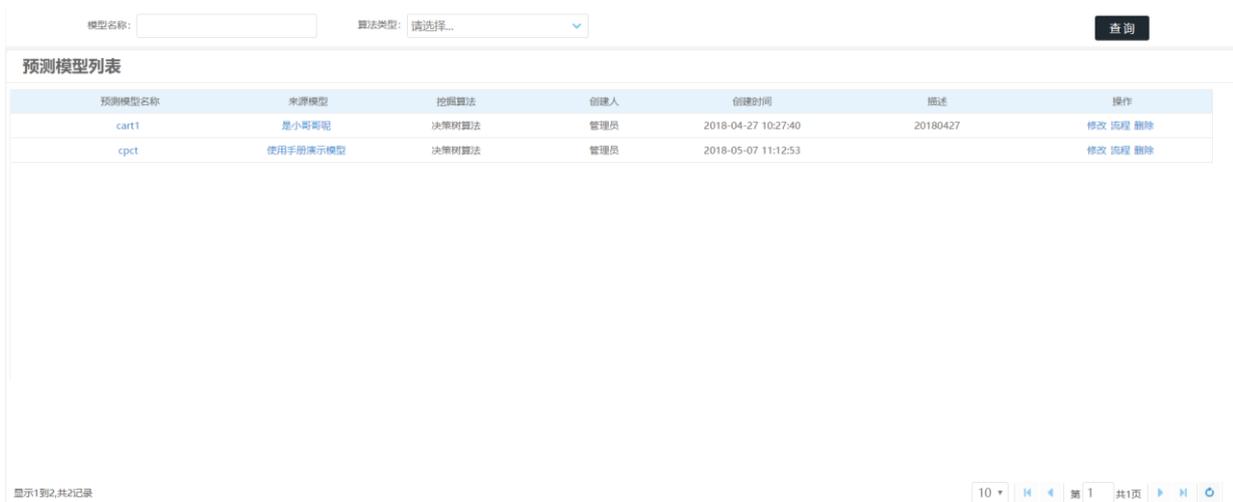
3.5.1.1 业务描述

对应菜单栏——流程开发-预测模型管理。

页面上方区域支持预测模型搜索。

页面中间区域，以列表展示所创建的预测模型，包括预测模型名称、来源模型、挖掘算法、创建人、创建日期、描述。

3.5.1.2 功能界面



预测模型名称	来源模型	挖掘算法	创建人	创建时间	描述	操作
cart1	星小哥哥呢	决策树算法	管理员	2018-04-27 10:27:40	20180427	修改 流程 删除
cpct	使用手册演示模型	决策树算法	管理员	2018-05-07 11:12:53		修改 流程 删除

3.5.1.3 操作步骤

1 预测模型搜索

支持按模型名称、算法类型的单一条件或组合条件的搜索，输入搜索条件后，点击右侧的“搜索”按钮即可。

2 预测模型列表操作

在列表中每一个样例模型的右侧，可对预测模型执行修改、删除、查看（流程）的操作。

3.5.1.4 注意事项

3.5.2 样例库管理

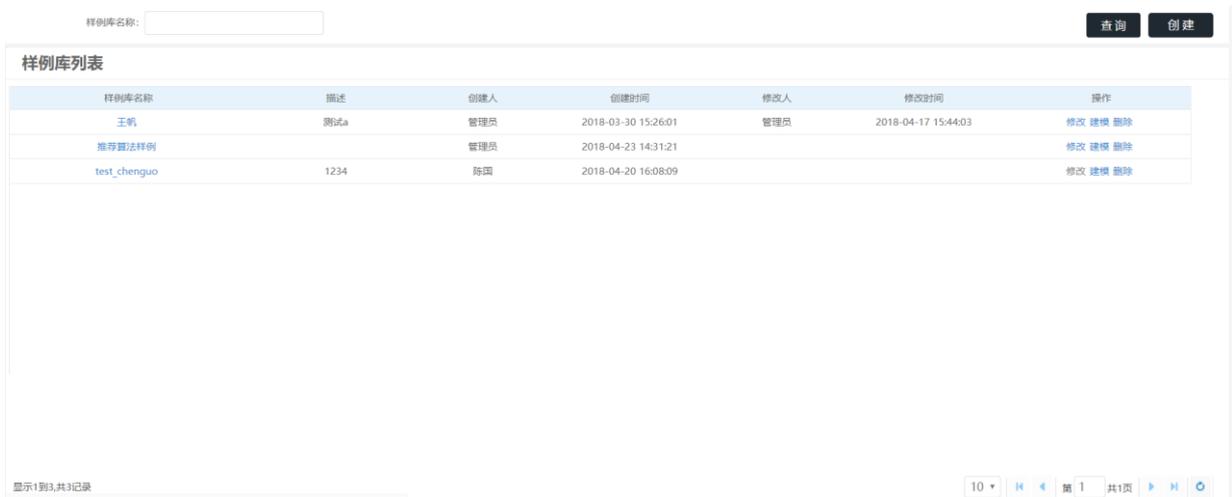
3.5.2.1 业务描述

对应菜单栏——流程开发-样例库管理。

页面上方区域支持样例库搜索。

页面中间区域，以列表展示所创建的样例库，包括样例库名称、修改人、修改时间、创建人、创建日期、描述。

3.5.2.2 功能界面



样例库名称	描述	创建人	创建时间	修改人	修改时间	操作
王帆	测试a	管理员	2018-03-30 15:26:01	管理员	2018-04-17 15:44:03	修改 建模 删除
推荐算法样例		管理员	2018-04-23 14:31:21			修改 建模 删除
test_chenguo	1234	陈国	2018-04-20 16:08:09			修改 建模 删除

3.5.2.3 操作步骤

1 样例库搜索

支持按样例库名称的搜索，输入搜索条件后，点击右侧的“搜索”按钮即可。

2 样例库列表操作

在列表中每一个样例模型的右侧，根据权限设置可对预测模型执行修改、删除、建模的操作。

3.5.2.4 注意事项

点击样例库名称即可查看样例流程。

3.5.3 缓存配置

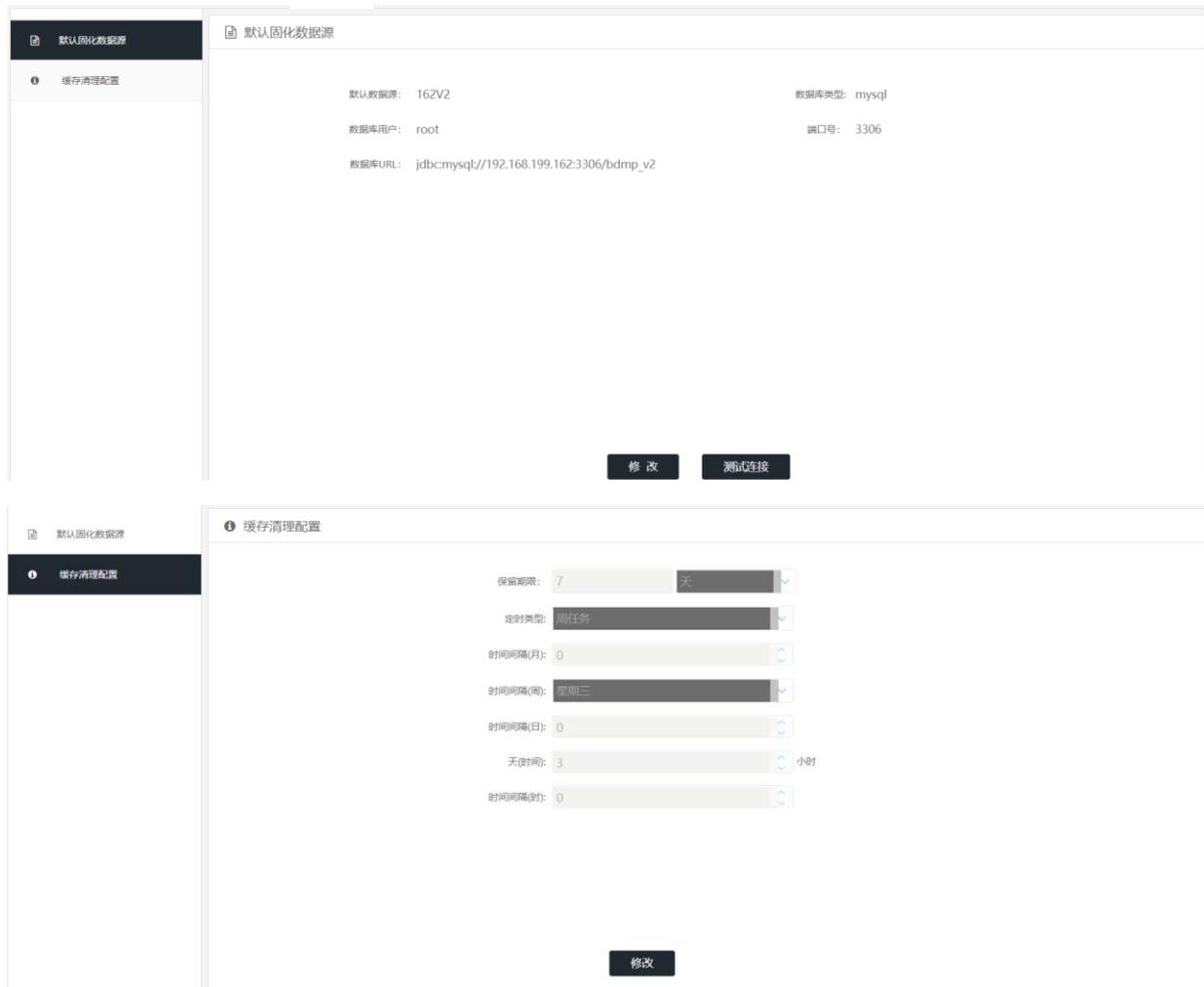
3.5.3.1 业务描述

对应菜单栏——流程开发-缓存配置。

页面左方区域支持默认固化数据源配置和缓存清理配置。

页面中间区域，以表格方式展现默认固化数据源的配置信息和缓存清理配置的详细信息。

3.5.3.2 功能界面



3.5.3.3 操作步骤

1 默认固化数据源配置

支持默认固化数据源的配置，点击修改后，即可选择相应的数据源。

2 缓存清理配置

支持缓存清理的配置，点击修改后，即可修改缓存清理的时间，定时清理缓存数据。

3.5.3.4 注意事项

缓存清理配置不会删除状态仍是固化的表。

3.6 权限说明

3.6.1.1 业务描述

用户归属于企业账号，用户可针对自己创建的模型，数据集，数据源，样例库等进行修改和查看及删除。超级管理员同样拥有这些权限。

3.6.1.2 功能界面

例

数据源管理							操作		
数据源名称	主机	端口	数据库名称	数据库类型	数据库驱动				
行云数据库	192.168.199.163	1801	BONC	xcloud	com.bonc.xcloud.jdbc.XCloudDriver				
oracle测试	192.168.199.170	1521	bfzcdb	oracle	oracle.jdbc.driver.OracleDriver				
基础研发部本地测试	192.168.199.175	3306	bdmp	mysql	com.mysql.jdbc.Driver				
162_bdmp_v1	192.168.199.162	3306	bdmp_v1	mysql	com.mysql.jdbc.Driver				
162V2	192.168.199.162	3306	bdmp_v2	mysql	com.mysql.jdbc.Driver				
王帆测试	127.0.0.1	3306	bdmp	mysql	com.mysql.jdbc.Driver				
欧晓呀	127.0.0.1	3306	bdmp_v2	mysql	com.mysql.jdbc.Driver				
hive数据源	192.168.199.170	10000	test	hive	org.apache.hive.jdbc.HiveDriver				
162Hive	192.168.199.162	10000	test	hive	org.apache.hive.jdbc.HiveDriver				
presto测试	192.168.199.170	18080	test	presto	com.facebook.presto.jdbc.PrestoDriver				

样例库列表							操作		
样例库名称	描述	创建人	创建时间	修改人	修改时间				
王帆	测试a	管理员	2018-03-30 15:26:01	管理员	2018-04-17 15:44:03				
推荐算法样例		管理员	2018-04-23 14:31:21						
test_chenguo	1234	陈国	2018-04-20 16:08:09						

3.6.1.3 注意事项

想将权限赋给他人，可通过用户授权或企业账号授权。

