**绿湾智慧数据自助建模平台**

**操作手册**

|  |  |
| --- | --- |
| **文档名称** | 操作手册 |
| **版本号** | V1.0 |
| **创建人** | 钟易庚 |
| **创建日期** | 2021/10/20 |
| **保密分类** | 内部公开 |

**目录**

[河池智慧公安信息化建设项目 1](#_Toc13723)

[1. 系统登陆 5](#_Toc13414)

[2. 模型超市 5](#_Toc25010)

[3. 模型列表 6](#_Toc18678)

[4. 建模画布 7](#_Toc16674)

[4.1. 画布工具栏 7](#_Toc7271)

[4.2. 数据和算子 8](#_Toc5677)

[4.3. 算子菜单栏 8](#_Toc16715)

[4.4. 数据预览区域 9](#_Toc1295)

[5. 算子介绍 10](#_Toc32540)

[5.1. 高级搜索类算子 10](#_Toc17118)

[5.1.1. 高级搜索算子 10](#_Toc7181)

[5.1.2. 时空搜索 10](#_Toc13454)

[5.1.3. 关系扩线算子 11](#_Toc2702)

[5.1.4. 组内关系算子 12](#_Toc18165)

[5.1.5. 以人搜事算子 13](#_Toc16311)

[5.1.6. 以事搜人算子 14](#_Toc20479)

[5.1.7. 身份关联算子 14](#_Toc25135)

[5.1.8. 通过号码找人 15](#_Toc2675)

[5.1.9. 文本搜索 15](#_Toc15707)

[5.2. 碰撞类算子 16](#_Toc2298)

[5.2.1. 数据求同 16](#_Toc713)

[5.2.2. 数据求合 17](#_Toc8775)

[5.2.3. 数据求差 17](#_Toc29753)

[5.3. 筛选类算子 18](#_Toc10128)

[5.3.1. 条件筛选 18](#_Toc8885)

[5.3.2. 对象筛选 18](#_Toc32695)

[5.3.3. 属性值筛选 19](#_Toc5008)

[5.4. 提取类算子 20](#_Toc21716)

[5.4.1. 提取表格一列 20](#_Toc25165)

[5.4.2. 提取对象属性 20](#_Toc12034)

[5.4.3. 提取文本字段 21](#_Toc28340)

[5.4.4. 二维表转对象 21](#_Toc28315)

[5.4.5. 对象转二维表 22](#_Toc24739)

[5.5. 数据挖掘类算子 22](#_Toc26948)

[5.5.1. 分组统计 22](#_Toc22369)

[5.6. 数据处理类算子 23](#_Toc18559)

[5.6.1. 字符截取 23](#_Toc2164)

[5.6.2. 数据排序 24](#_Toc32153)

[5.6.3. 数据去重 24](#_Toc29263)

[5.7. 可视化类算子 25](#_Toc32298)

[5.7.1. 图谱可视化 25](#_Toc10583)

[5.7.2. 图表可视化 26](#_Toc21893)

[5.7.3. 地图可视化 26](#_Toc14854)

[5.8. 数据上传算子 27](#_Toc15057)

[5.8.1. 上传文件 27](#_Toc12602)

# 系统登陆

用户可以通过账号密码登陆，或者PKI登陆两种方式登陆系统。

其中，账号密码登陆时，用户可以选择身份证号或警号作为账号，进行系统登陆。身份证号或警号用于进行用户的身份校验。

如果采用账号密码登陆形式，请注意每个月更换密码以确保账号安全。



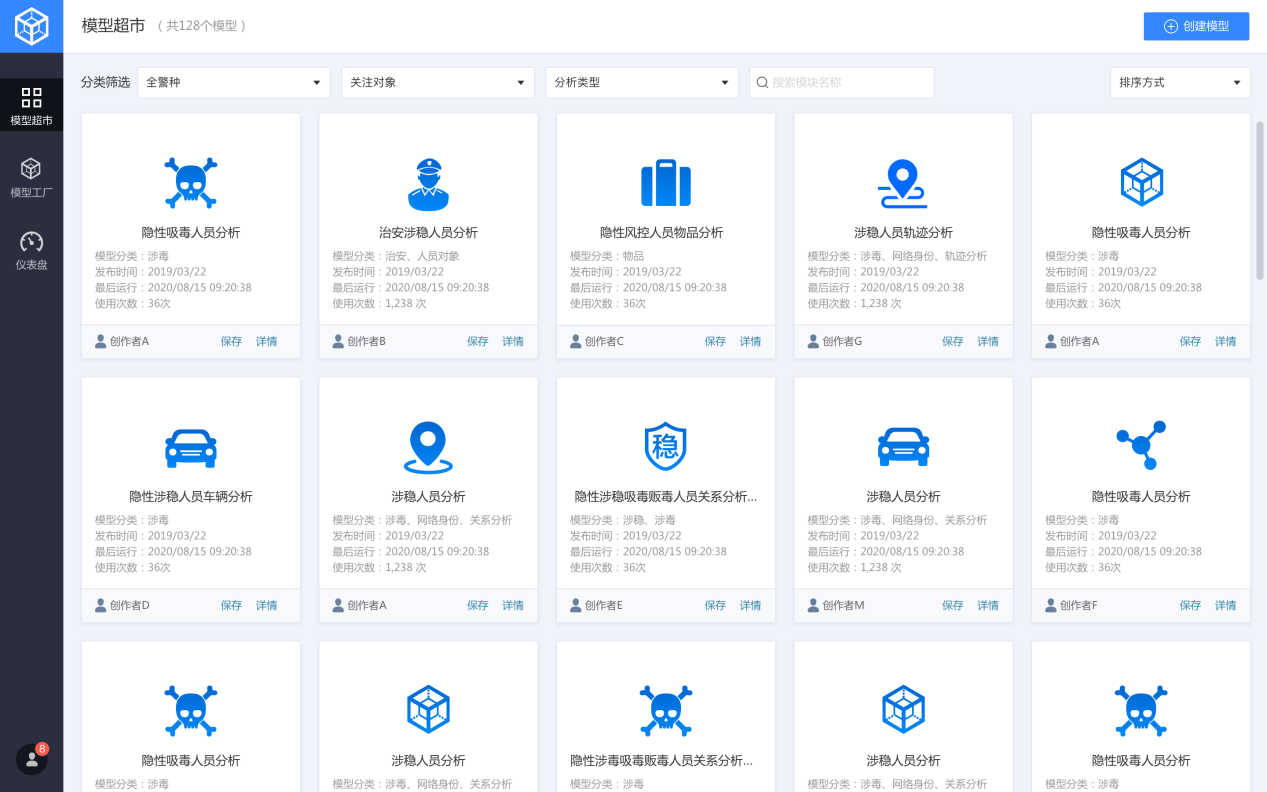
# 模型超市

用户进入自助建模系统后，首先会进入模型超市页面。模型超市页面以卡片形式展示了目前已在模型超市中发布的所有模型。用户可以根据模型的分类对模型进行筛选，或根据模型名称对模型进行搜索，还可以按照模型的发布时间、热度等对模型进行排序。

每个模型卡片右下角都有“详情”和“保存”按钮。

点击详情按钮，用户可以跳转至画布页面查看模型的搭建逻辑、参数配置以及结果数据（以预览的形式）。在模型超市中的模型，在画布下都是只读的，不可编辑。

点击保存按钮，用户可以将该模型保存至其自己的模型列表中，对该模型进行后续的编辑与使用。



# 模型列表

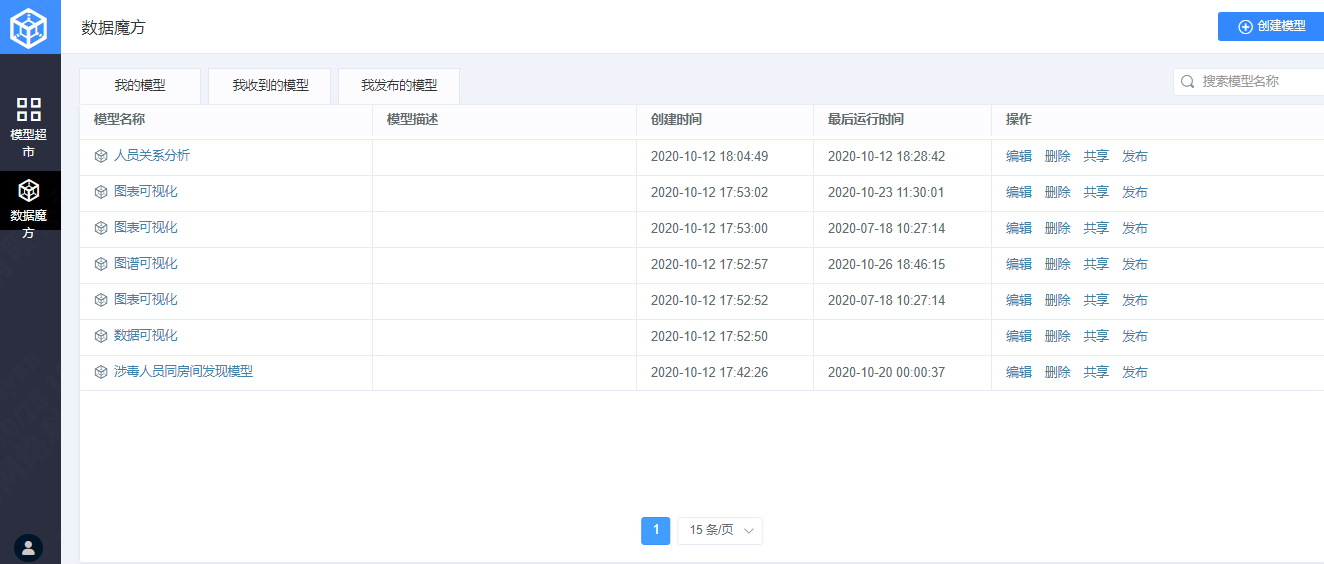
模型列表分为：我的模型列表，我收到的模型列表，我发布的模型列表。其中：

我的模型列表中的所有模型，都是用户可编辑的，并且可以进行发布、共享、删除操作。

发布会将该模型发布到模型超市，同时会在“我发布的模型列表”中备份该模型。用户可以在“我发布的模型列表”中，维护自己已发布的模型，对其进行下架或删除操作。（如果模型在发布状态被删除，则模型超市会同步下架该模型）

共享可以将“我的模型列表”中的模型，点对点的共享给某个或某些用户。共享后的模型会出现在对方的“我收到的模型列表”中。用户可以通过在“我收到的模型列表”中对模型进行保存，将其保存至“我的模型列表”，进行后续的编辑或使用。

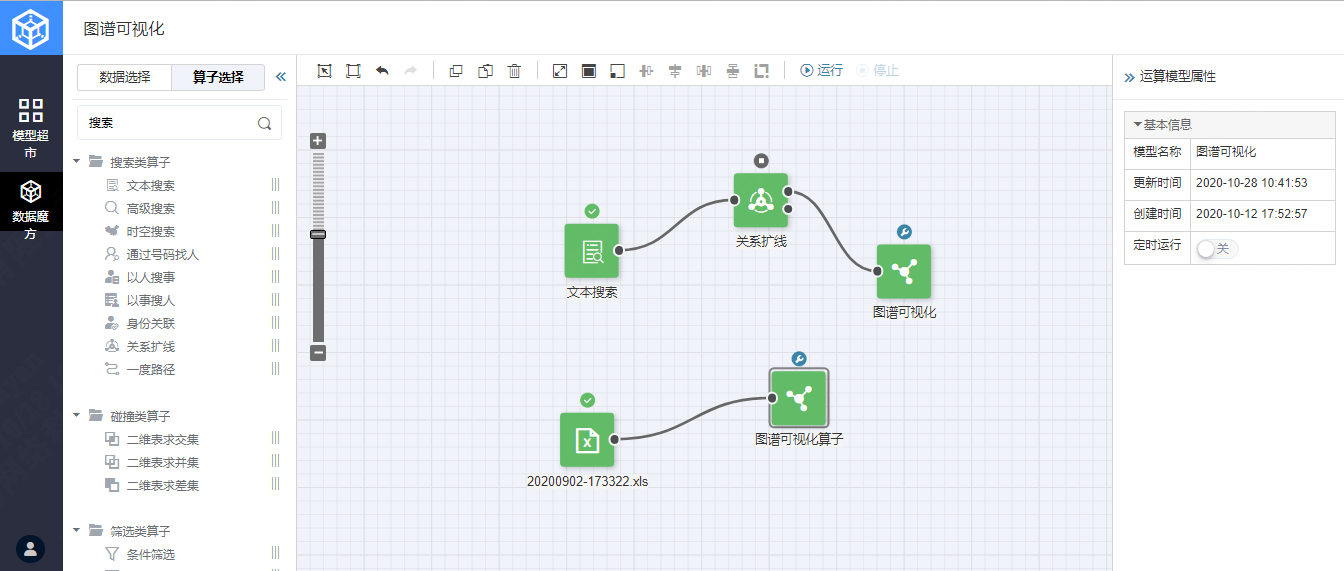
删除会将当前模型删除，删除操作不可恢复。



# 建模画布

当前为自助建模系统建模画布主界面。其中心区域为建模画布，左侧为“数据&算子选择栏”、顶部为“画布工具栏”、右侧为属性栏。

建模画布是用户进行建模的主区域。用户可以从左侧的“数据&算子选择栏”将需要用到的数据与算子，拖拽至建模画布区域，并将算子与数据相互之间连线，来构建模型。



## 画布工具栏

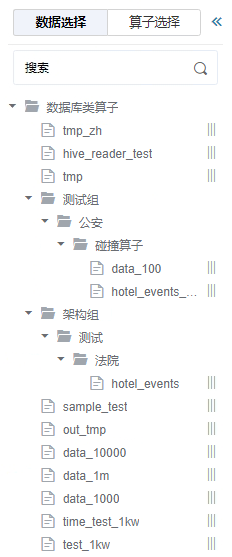
当前为画布工具栏，用于对画布上的算子进行操作。操作包括选中、反选、撤销、复制、粘贴、删除、按不同方式对算子进行排列等。



## **数据和算子**

用户可以使用位于左侧的“数据&算子选择栏”来选择建模需要用到的数据与算子。数据与算子的选择以顶部选项卡的形式进行切换。

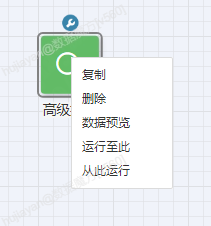
整个栏位顶部的搜索框可以用于对整个选择栏进行搜索。

## **算子菜单栏**

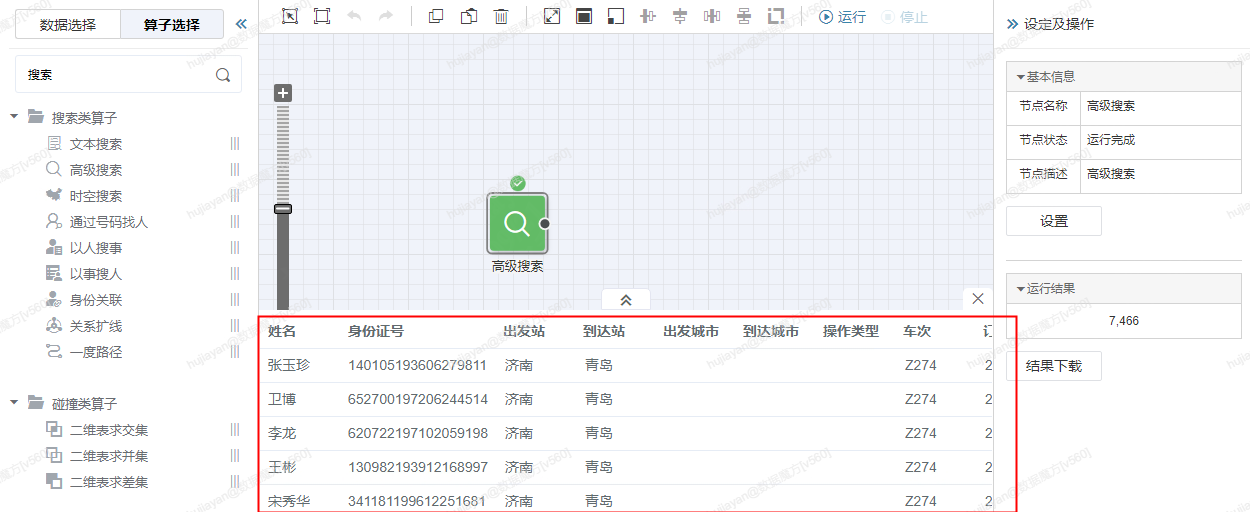
用户可以通过在画布中的算子上，点击右键的方式，唤出算子菜单栏。算子菜单栏包含操作：

1. 复制：复制当前算子
2. 删除：删除当前算子
3. 数据预览（只有当前算子配置完成时可用）：查看当前算子的样例数据（100条）
4. 运行至此：从可以连接到的最左侧节点一直运行至当前节点
5. 从此运行：从当前节点一直运行至可以连接到的最末端节点



## **数据预览区域**

当对算子进行数据预览后，会在主页面下方弹出数据预览区域，展示该次数据预览的样例数据。该数据预览区域可以进行收放或关闭。



# 算子介绍

## 高级搜索类算子

### 高级搜索算子

根据指定的搜索对象及其属性，在融合数据中查找对应数据。

输入：无

输出：对象（ROWKEY）

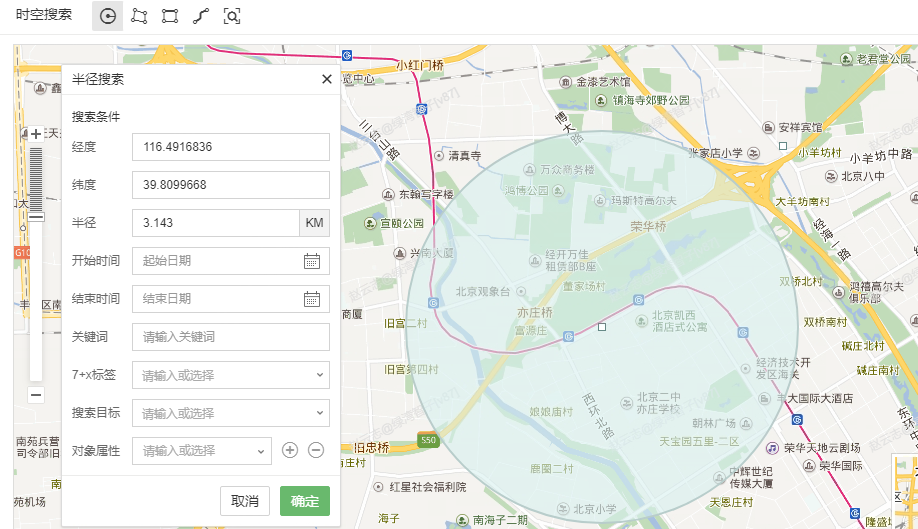


### **时空搜索**

在地图上通过指定时间、空间和属性条件，对融合数据进行搜索。

输入：无

输出：对象（ROWKEY）



### **关系扩线算子**

根据指定的输入（对象），按不同关系类型进行扩线，输出对应包含身份证号或手机号的扩线后的数据。

输入：对象

输出：对象或二维表

与人（身份证号）相关的关系包括：直接关系、疑似关系、同酒店、同网吧、同航班、同火车、同核录、同案件、通话、快递、银行转账、属性匹配。

与手机相关的关系包括：通话关系、快递关系、属性匹配。

可选择全部关系，此时将会根据选择的对象类型，自动匹配可搜索的关系。



### **组内关系算子**

根据指定的输入（对象），按不同关系类型，在输入对象间搜索符合搜索条件的数据。

输入：对象

输出：对象或二维表

与人（身份证号）相关的关系包括：直接关系、疑似关系、同酒店、同网吧、同航班、同火车、同核录、同案件、通话、快递、银行转账、属性匹配。

与手机相关的关系包括：通话关系、快递关系、属性匹配。

可选择全部关系，此时将会根据选择的对象类型，自动匹配可搜索的关系。



### **以人搜事算子**

查询人对象所关联的，符合相应搜索条件的所有事件信息。

输入：对象（身份证号、手机号、车牌号、银行卡）

输出：对象

身份证号可关联到所有事件。

手机号可关联到：通话事件、快递事件、手机定位事件、号址关系记录、电话号码开户事件、手机号注册信息。

车牌号可关联到：卡口过车、车辆定位、车辆盘查、停车记录、违章记录。

银行卡可关联到：银行卡转账记录、银行卡余额、银行卡冻结事件。

可选择全部事件，此时将会根据输入的对象类型，自动匹配可搜索的事件。

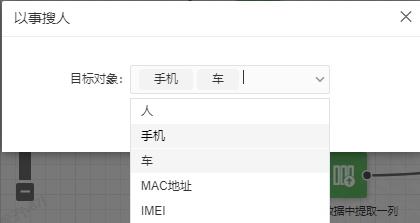


### **以事搜人算子**

根据输入的事件对象与所选的目标对象类型，输出目标对象。

输入：对象

输出：对象或二维表



### **身份关联算子**

基于输入的对象，搜索某一种其所关联的所有身份。

输入：对象

输出：对象或二维表

人可以关联到的身份有：手机、车、mac、imei、qq、微信、虚拟帐号。

手机可以关联到的身份有：人、mac、imei、qq、微信、虚拟帐号。

车可以关联到的身份有：人。

mac、imei、qq、微信、虚拟帐号可以关联到的身份有：人、手机。



### **通过号码找人**

通过身份证号或者手机号，找到对应的人员信息

输入：一列省份证号或手机号

输出：对象



## **碰撞类算子**

### 数据求同

基于选择的字段，对两张二维表进行交运算。

输入：二维表

输出：二维表

在输入的二维表中，选择一组或多组字段，进行交运算。字段间可以是满足任意条件，或满足全部条件。同时，可以根据两张输入二维表中的字段，来生成输出二维表的字段。



### **数据求合**

基于选择的字段，对两张二维表进行并运算。

输入：二维表

输出：二维表

在输入的二维表中，选择一组或多组字段，进行并运算。可以选择是否去重，如果选择去重，则将所选字段下的数据合并，然后去掉完全相同项；如果选择不去重，则直接将所有已选数据合并。输出的二维表字段，同并运算的字段。



### **数据求差**

基于选择的字段，对两张二维表进行差运算。

输入：二维表

输出：二维表

在输入的二维表中，选择一组或多组字段，进行差运算（默认数据集1差数据集2）。如果选择了多个数据项，则将多个数据项的数据拼接起来，再做差运算。差算子可以根据两张输入二维表中的字段，来生成输出二维表的字段。



## **筛选类算子**

### 条件筛选

根据条件对二维表中的数据进行筛选

输入：二维表

输出：二维表

在输入的二维表中，选择一个或多个字段，并针对每个选择的字段，选择其筛选条件。多个字段的筛选条件间，可以是满足任意关系或满足全部关系。输出满足筛选条件的所有数据，输出二维表与输入二维表的字段相同。



### **对象筛选**

根据指定的条件，对输入的对象进行筛选，输出满足搜索条件的数据。

输入：对象

输出：对象

在输入的对象中，选择一个或多个属性，并针对每个选择的属性，选择其筛选条件。多个属性的筛选条件间，可以是满足任意关系或满足全部关系。输出满足筛选条件的所有对象。



### **属性值筛选**

通过一列数据，对已有的对象的某个属性进行筛选，得到符合条件的对象数据。例如，通过一列身份证号，对一批人进行筛选，得到身份证号在这列数据中的所有人员信息。

输入：一列数据，对象（ROWKEY）

输出：对象（ROWKEY）



## **数据挖掘类算子**

### 分组统计

将二维表中的数据以某一维度分组，并对分组后的数据进行统计。

输入：二维表

输出：二维表

统计方式包括：计数、求和、求平均值、去重计数（将统计字段去重后再计数）、最大N个，最小N个，最早N个，最晚N个。其中，当统计方式被选择为计数、求和、求平均值、去重计数时，可以选择多个统计字段。系统会自动识别所选择的统计字段的类型，并匹配可选的统计方式供用户选择。



## **数据处理类算子**

### 字符截取

按位数对二维表中的数据进行截取。

输入：二维表

输出：二维表

1. 用户在截取字段处选择需要操作的表字段（单选）。
2. 用户选择要截取的字段的位数，其中位数可以正数也可也倒数
3. 用户制定截取后生成的新字段的名称
4. 点击确定，系统自动根据用户设置的位数，对整列数据进行截取，截取结果保存在新建字段中。
5. 如果位数超过最大位数（例如截取正数1-正数20，但某条数据只有19位，则截取1-19位）
6. 如果截取位数小于1，则输出空字段。



### **数据排序**

对二维表数据进行排序。

输入：二维表

输出：二维表

1. 用户需要指定排序字段，排序字段可多选。
2. 排序时，优先级按照排序字段从上到下的顺序进行排序。
3. 每个排序字段，都支持以单独的升序或降序方式进行排序。



### **数据去重**

对二维表数据进行去重。

输入：二维表

输出：二维表

1. 去重字段下拉框中可以多选，也可以直接勾选全部。
2. 该算子的输出的表字段，为用户勾选的去重字段。



## **可视化类算子**

### 图谱可视化

对接入的对象数据或二维表数据，进行图谱形式的展示。

如果输入为对象数据（前置算子为组内关系，关联扩线、身份关联）时，不需要进行配置，自动生成关系图谱并展示；

如果输入为二维表数据时，用户选择需要关联的字段项，并且选择关系类型（支持自定义），然后系统根据用户的配置生成关系图谱并展示。

输入：对象或二维表（二选一）

输出：图谱



### **图表可视化**

将接入的二维表数据进行图表可视化展示。目前支持折线图、柱状图、饼图。

输入：二维表

输出：图表



### **地图可视化**

将接入的二维表数据或对象数据进行地图可视化展示。目前支持聚合散点图和热力图两种地图展示形式。

当前置输入为对象数据时，无需尽心配置，系统自动使用对象数据中的地理位置信息进行上图展示；

当前置输入为二维表数据时，可以选择根据经纬度字段上图或根据地址信息字段上图。其中，使用经纬度上图时，需指定对应的坐标系。

输入：对象或二维表（二选一）

输出：地图



## **数据上传算子**

### 上传文件

将excel或.csv文件，作为数据源上传至建模平台，参与模型运算。

输入：文件

输出：二维表

