

V4金智塔隐私计算平台产品功能介绍

金智塔隐私计算平台

产品功能介绍

前言

修改记录

文档版本	修改说明	发布日期	作者	签发
V1.0	初版	2023.02.20	龚松挺、沈阳、 潘韦绮、张豹、 沈琦、周凯明、 王克华、丰佩、 朱明杰	巫锡斌

1. 平台简介

金智塔隐私计算平台在国家重点研发项目（NO.2018YFB1403001）支持下，融合了多方安全计算、联邦学习、区块链、数字水印等技术，提供了数据确权、分级分类、质量审计、应用存证等能力，实现了数据可用不可见、用途可控可计量，具有高安全、高性能、高扩展、高互通四大优势，已通过中国信通院可信隐私计算基础能力、性能、安全系列评测，央行国家金融科技检测

中心的「多方安全计算金融应用评测」，公安部安全认证，以及华为鲲鹏认证等。平台已在数字政务、智慧金融、智慧产业等场景打造了数十个成功应用案例，其数字政务、智慧金融领域的应用实践分别入选了中国信通院2022大数据「星河（Galaxy）」奖项的「标杆案例」和「优秀案例」。

金智塔隐私计算平台构建了去中心全对称分布式架构，提高了网络的可扩展性、节点的可复用性；支持混合协议联合SQL，极大降低了隐私计算数据分析门槛；采用了大规模、分布式并行架构和内置多模态异构计算引擎，支持亿级隐私求交、千万级联合建模；并以“算法插件”形式接入不同平台，实现了隐私计算平台间的互联互通；且支持信创国产化，全面保障数据价值安全释放。



图1 平台简介

金智塔隐私计算平台基于联邦学习和隐私计算两大技术路线，支持多节点数据联合计算的同时，亦支持单节点数据安全计算。目前已在风控、营销、监管等多个金融场景中实践应用。

平台特点如下：

- 高性能：平台基于分层设计理念，可基于不同的业务场景做算法组合定制，实现性能稳定和计算效率的平衡；
- 易扩展：全对称的分布式架构，节点扩展成本低。网状结构，节点扩展数量不受限；
- 高安全：原语层基于不经意传输、秘密分享、同态加密等密码学技术，计算通讯信息支持区块链全程存证溯源；
- 易互通：针对“计算孤岛”问题，平台开放系统层接口，已实现与部分金融机构和数源单位的互联互通。

2. 架构设计

2.1. 总体架构

金智塔隐私计算平台总体架构如下图2所示，自底向上分为：部署适配层、计算引擎层、计算平台层。

部署适配层用于适配不同的软硬件环境，屏蔽基础设施差异，向上层提供统一的运行时接口。集中式服务部署于数据中心，其执行环境如计算、存储、网络等是同构的；而隐私计算服务需要部署在不同的参与机构，各家的基础设施不尽相同，面对着复杂的异构执行环境。例如，在计算资源上，有的机构采用x86架构服务器，有的机构采用ARM架构服务器；在存储资源上，各个机构提供的数据库产品（MySQL、Oracle、DB2等）、对象存储、文件存储存在较大差异；在网络拓扑上，各个机构的网络设施错综复杂，跨机构的网络链路较长，存在防火墙、正向/反向代理、4/7层负载均衡等各种网络设备。为了解决以上问题，部署适配层以容器化技术为核心，屏蔽计算资源差异，支持Docker、Kubernetes等部署方式；通过存储资源抽象，提供统一的IO接口，屏蔽存储资源差异；通过域名解析，统一机构内和机构间服务访问模式，屏蔽网络链路差异。

计算引擎层实现了隐私计算核心的协议和算法，分为安全原语和安全算法两层。安全原语层提供了隐私计算核心的密码学协议实现，包括多方安全计算、同态加密、差分隐私、不经意传输等，提供了安全的数值运算、比较运算、矩阵运算等计算原语。安全算法层在此基础上，实现了用于数据分析和建模的各类算法，如隐私求交、匿踪查询、联合建模、联合SQL等。安全算法层同时提供了基于多方安全计算和基于联邦学习两大技术路线的算法实现，用户可以根据业务、数据量、安全要求等因素，选择合适安全等级的隐私计算算法，以平衡算法的安全和性能。

计算平台层提供了可视化的操作界面，分为管理台和工作台两部分。金智塔隐私计算平台是一个多租户平台，管理台面向平台运营人员，提供了组织、人员、角色、权限等多租户管理能力。同时，管理台提供了节点注册、节点授权、节点组网等多方项目协同管理能力。此外，管理台还提供了对平台进行日志审计、用户行为审计、项目审计等多维度运行监管能力，保障平台数据安全。工作台面向普通建模人员，以拖拉拽形式提供了可视化建模工具。用户以DAG形式构建工作流，工作台调度器负责多方任务协同调度，执行器负责任任务的全生命周期管理。

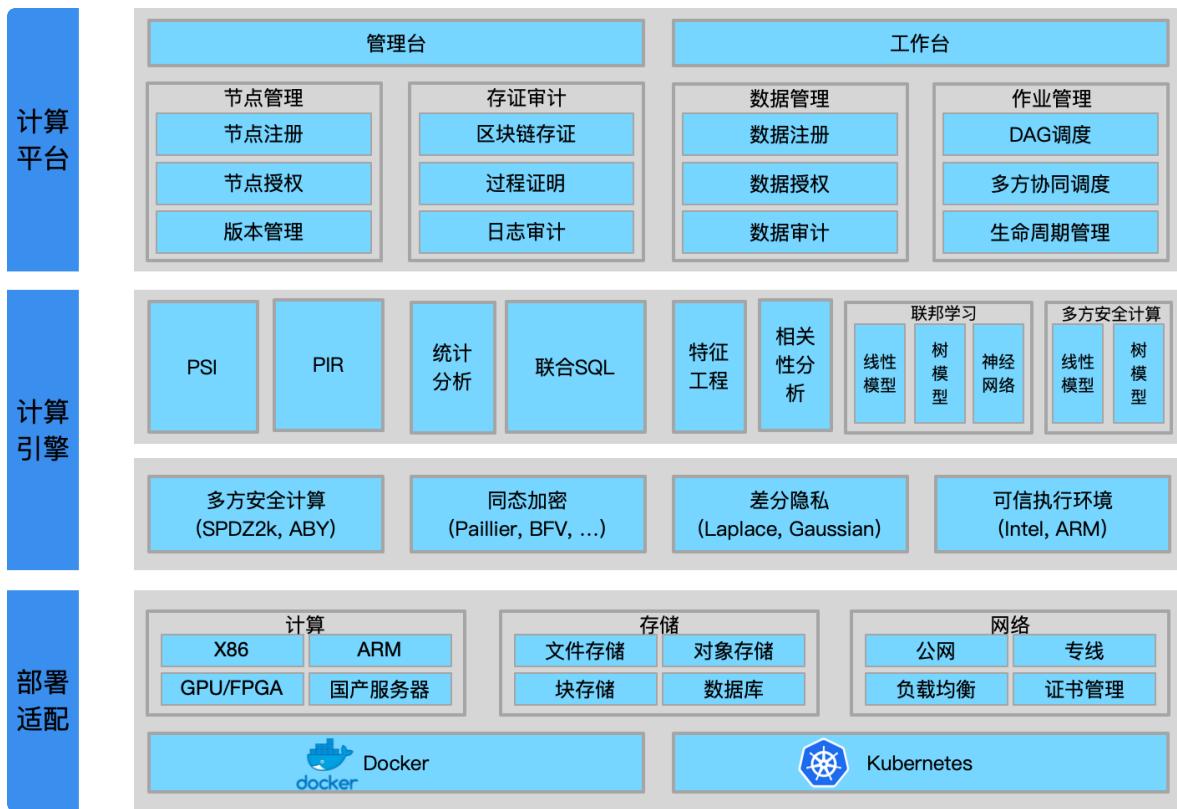


图2 系统架构

2.2. 部署架构

隐私计算平台由多个隐私计算节点组成。各隐私计算节点均可提供任务计算和数据提供的功能，同时各隐私计算节点之间需要根据具体的合作协议进行网络的连通，如若不需要进行数据合作，则不需要进行网络的连通。

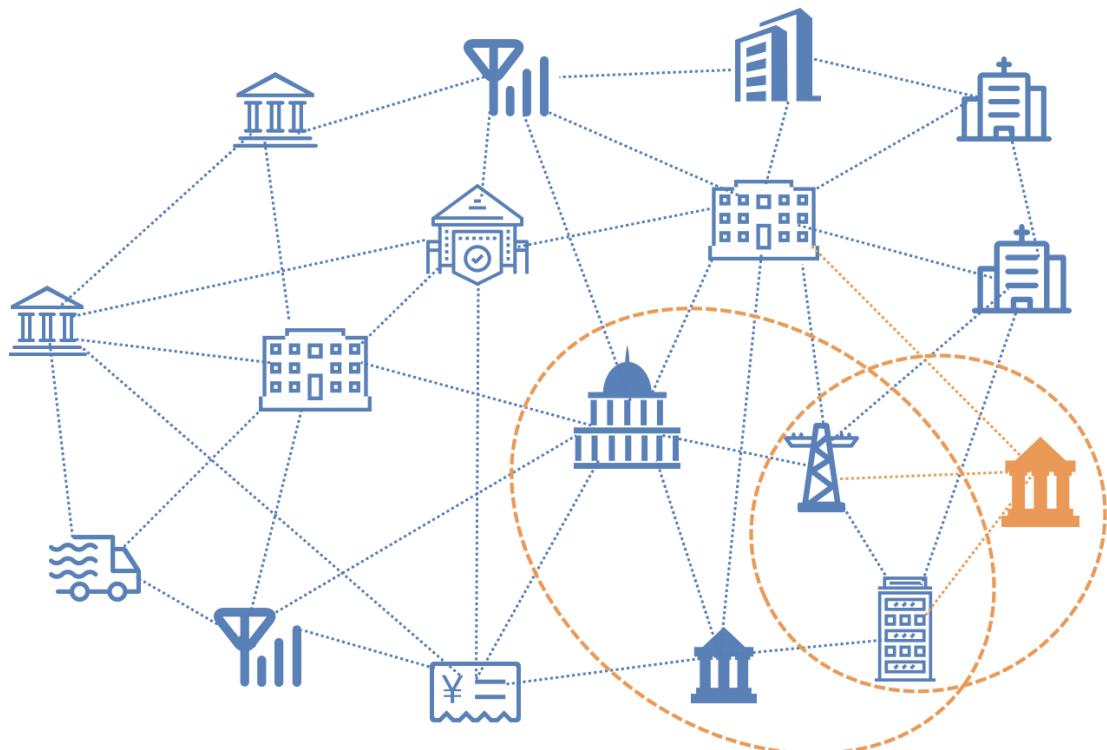


图3 部署架构

2.3. 隐私计算流程

通过安全的算法和协议，参与方将明文形式的数据加密后或转化后再提供给其他方，任一参与方都无法接触到其他方的明文形式的数据，从而保证各方数据的安全。隐私计算平台由各个隐私计算节点组成，其执行流程如下图所示：

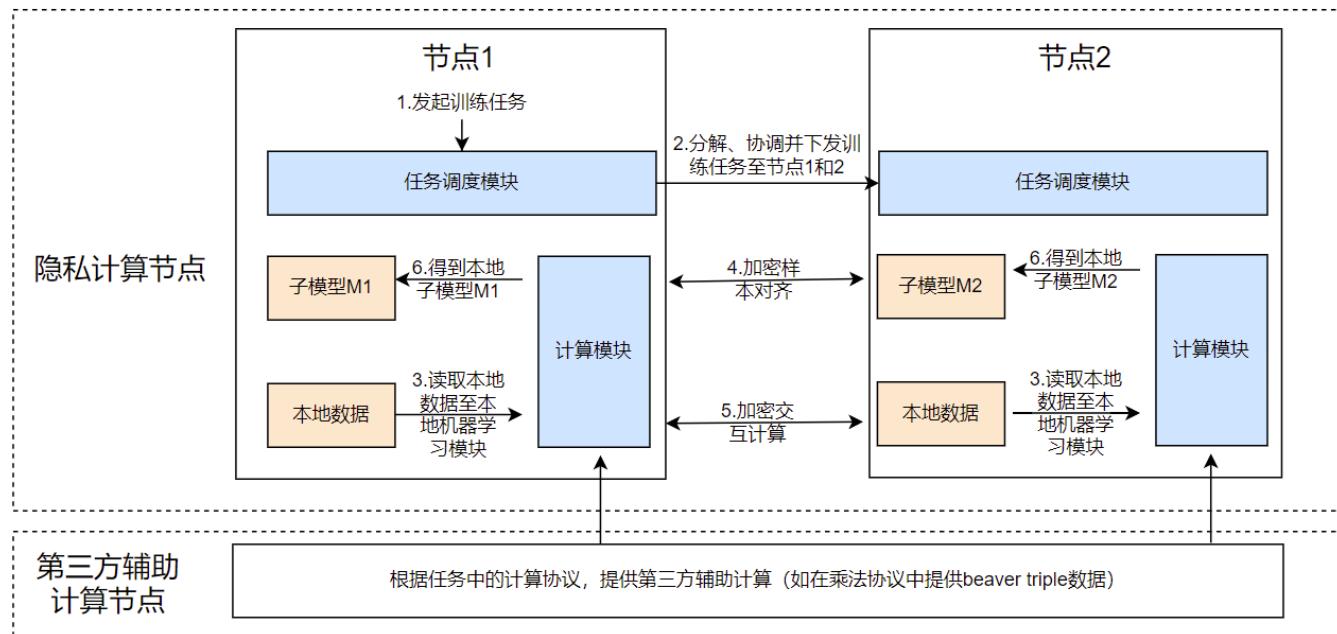


图4 隐私计算流程

3.管理台

3.1.通讯设置

3.1.1. 节点列表

在节点列表页面可以进行新增节点的操作。导入节点证书点击确认添加成功。此时，管理台获取新增节点如下信息：节点名称、节点ID，公钥信息。同时也可以支持将某节点退出。

The screenshot shows the 'Node List' page of the 'Communication Settings' section. On the left sidebar, 'Node List' is selected. The main area displays a table with columns: 'Node Name', 'Node ID', 'Public Key', and 'Operations'. Two nodes are listed: 'Bob' with ID 'bob' and 'Public Key' status '已获取' (Acquired), and 'Carl' with ID 'carl' and 'Public Key' status '已获取' (Acquired). A blue button labeled 'Add Node' is visible at the top left of the table area.

节点名称	节点ID	公钥	操作
鲍勃	bob	已获取	<button>删除</button>
卡尔	carl	已获取	<button>删除</button>

图5 节点列表

点击“新增节点”按钮，通过上传文件或者输入具体的节点证书编码添加节点。（节点证书由其他节点提供）

The screenshot shows the 'Add Node' dialog box overlaid on the node list page. The dialog has a red border and contains fields for 'Node Name' (selected 'Import Certificate') and 'Public Key'. It includes a file upload input field labeled 'Import Node Certificate' and a 'Upload File' button. A red arrow points from the 'Add Node' button on the main page to this dialog box.

图6 新增节点

点击需要删除的节点右侧“删除”按钮，删除节点

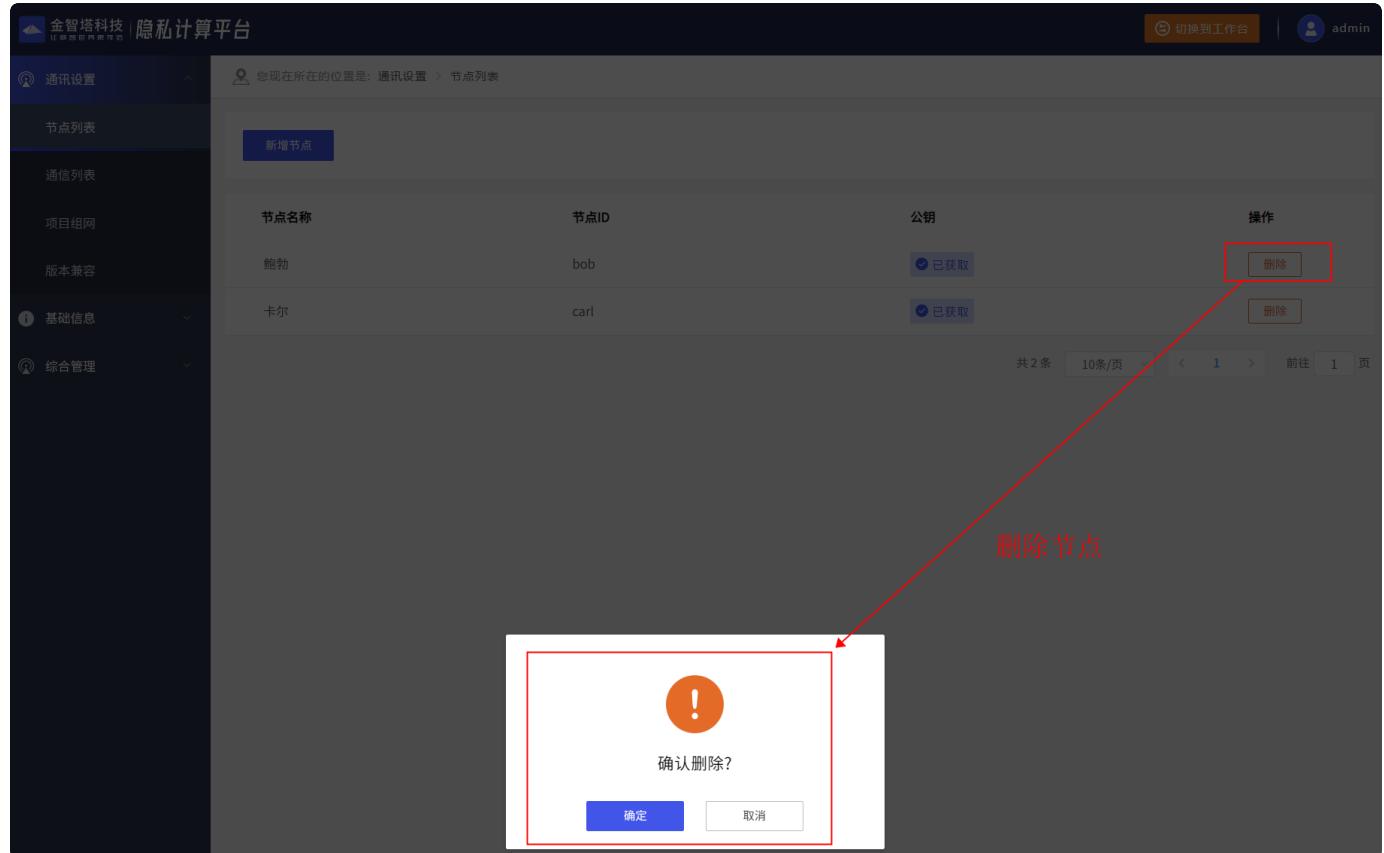


图7 节点删除

3.1.2. 通信列表

在通信列表中可以完成本节点与其他节点的通信操作。点击发起通信选择接收节点，选择通信形式并填入通信地址，待接收节点批复授权互通，即可完成本节点到接收节点的通信。接收节点在通信完成后支持随时废止通信。

通信列表中，展示了所有的网络链路，其中上方展示通信的图形化关系，下方展示具体记录。

点击“发起通信”按钮，选择接收节点，通讯形式，并输入目标地址`http://ip:port`。（需提前完成[新增节点](#)）

The screenshot shows the 'Communication Settings' section of the platform. On the left sidebar, under 'Communication List', there is a red box around the 'Initiate Communication' button. Below this, a table lists two communication records:

编号	发送节点	接收节点
112	alice	carl
110	alice	bob

The main area displays two nodes, 'carl' and 'bob'. Node 'carl' has a red box around its 'Initiate Communication' button. A red arrow points from this button to a callout box containing the following text:

1. 点击发起通信
2. 选择接受节点
3. 输入目标地址 (ip和port以具体部署时的网络配置为准)

A modal window titled 'Initiate Communication' is shown, also with a red box around it. It contains the following fields:

* 接收节点	鲍勃
* 通讯形式	直连
* 目标地址	http://192.168.88.102:10089

At the bottom of the modal are '完成' (Finish) and '取消' (Cancel) buttons.

图8 作为发送方，发起通信

作为发送方，可查看所有网络链路申请记录的详情和当前链路状态。

金智塔科技 隐私计算平台

切换到工作台 | admin

通信设置

节点列表

通信列表

项目组网

版本兼容

基础信息

综合管理

您现在所在的位置是: 通信设置 > 通信列表

发起通信

作为发送方

作为接收方

可拖拽、ctrl+滚轮缩放查看

可拖拽、ctrl+滚轮缩放查看

编号	发送节点	接收节点	时间	是否直连	目标地址	中转地址	申请状态	通信状态	操作
112	alice	carl	2023-01-12 20:28:48	路由	--	bob	已同意	断开	废止
110	alice	bob	2023-01-12 20:27:50	直连	http://192.168.8.102:10089	--	未批复	断开	重发

共 2 条 10条/页 < 1 > 前往 1 页

4. 新增的网络链路的详情

- 申请状态: 未批复
- 通信状态: 断开

图9 作为发送方，查看新增的链路记录

金智塔科技 | 隐私计算平台

切换到工作台 | admin

通讯设置

节点列表

通信列表

项目组网

版本兼容

基础信息

综合管理

您现在所在的位置是: 通讯设置 > 通信列表

发起通信

作为发送方

作为接收方

作为发送方

作为接收方

编号	发送节点	接收节点	时间	是否直连	目标地址	中转地址	申请状态	通信状态	操作
112	alice	carl	2023-01-12 20:28:48	路由	--	bob	<input checked="" type="checkbox"/> 已同意	<input checked="" type="checkbox"/> 正常	<button>废止</button>
110	alice	bob	2023-01-12 20:27:50	直连	http://192.168.8.8.102:10089	--	<input checked="" type="checkbox"/> 已同意	<input checked="" type="checkbox"/> 正常	<button>废止</button>

5. 接受节点在同意此条链路后，状态更新为
 • 申请状态：已同意
 • 通信状态：正常

共 2 条 10条/页 < 1 > 前往 1 页

图10 作为发送方，查看链路授权通过

作为发送方，可以对已经授权通过的链路进行废止，点击废止后，状态如下图所示。

金智塔科技 | 隐私计算平台

切换到工作台 | admin

通讯设置

节点列表

通信列表

项目组网

版本兼容

基础信息

综合管理

您现在所在的位置是: 通讯设置 > 通信列表

发起通信

作为发送方

作为接收方

作为发送方

作为接收方

编号	发送节点	接收节点	时间	是否直连	目标地址	中转地址	申请状态	通信状态	操作
112	alice	carl	2023-01-12 20:28:48	路由	--	bob	<input checked="" type="checkbox"/> 已同意	<input checked="" type="checkbox"/> 正常	<button>废止</button>
110	alice	bob	2023-01-12 20:27:50	直连	http://192.168.8.8.102:10089	--	<input type="checkbox"/> 废止	<input checked="" type="checkbox"/> 断开	

6. 作为发送方，废止链路后的状态如下：
 • 申请状态：废止
 • 通信状态：断开

图11 作为发起方，对链路进行废止

作为接收方，可对其他机构提交的发起通信互通请求进行批复，同时可查看历史批复记录和网络链路当前的状态。

金智塔科技 | 隐私计算平台

切换到工作台 | admin

您现在所在的位置是: 通讯设置 > 通信列表

发起通信

作为发送方 carl 爱丽丝 本节点 断开

作为接收方 bob 爱丽丝 本节点 断开

基础信息

综合管理

1. 作为接收方，可以看到申请的该条链路，状态如下：

- 申请状态：未批复
- 通信状态：断开

编号	发送节点	接收节点	时间	是否直连	目标地址	中转地址	申请状态	通信状态	操作
113	carl	alice	2023-01-12 20:30:39	直连	--	--	未批复	断开	批复
111	bob	alice	2023-01-12 20:27:58	直连	--	--	未批复	断开	批复

共 2 条 10条/页 < 1 > 前往 1 页

图12 作为接受方，查看合作方的申请通信记录

作为接收方，可点击“批复”按钮，对节点授权进行批复或拒绝互通。

作为发送方

carl

断开

爱丽丝
本节点

作为接收方

bob

断开

爱丽丝
本节点

carl

断开

爱丽丝
本节点

批复

授权互通 拒绝互通

确定 取消

2. 作为接收方，可以点击批复，进行授权互通或拒绝互通的操作

编号	发送节点	接收节点
113	carl	alice
111	bob	alice

地址 申请状态 操作

未批复 断开 批复

未批复 断开 批复

共 2 条 10 条/页 前往 1 页

图13 作为接收方，对申请通信的链路进行授权操作

作为发送方

carl

断开

爱丽丝
本节点

作为接收方

bob

断开

爱丽丝
本节点

carl

断开

爱丽丝
本节点

批复

授权互通 拒绝互通

确定 取消

3. 作为接收方，点击授权互通后，状态如下：

- 申请状态：已同意
- 通信状态：断开（需等待2min后才刷新）

编号	发送节点	接收节点	时间	是否直连	目标地址	中转地址	申请状态	通信状态	操作
113	carl	alice	2023-01-12 20:30:39	直连	--	--	未批复	断开	批复
111	bob	alice	2023-01-12 20:27:58	直连	--	--	已同意	断开	废止

共 2 条 10 条/页 前往 1 页

图14 作为接收方，对某条链路授权互通

The screenshot shows the 'Communication Settings' section of the platform. On the left sidebar, under 'Communication Settings', the 'Communication List' option is selected. In the main area, there are two diagrams illustrating the communication process:

- Left Diagram (As Sender):** Shows a connection between 'carl' and '爱丽丝 本节点'. A red dashed arrow labeled '断开' (Break) indicates a disconnection.
- Right Diagram (As Receiver):** Shows a connection between 'bob' and '爱丽丝 本节点'. A solid blue arrow labeled '正常' (Normal) indicates a normal connection.

Below the diagrams is a table of communication records:

编号	发送节点	接收节点	时间	是否直连	目标地址	中转地址	申请状态	通信状态	操作
113	carl	alice	2023-01-12 20:30:39	直连	--	--	未批复	断开	批复
111	bob	alice	2023-01-12 20:27:58	直连	--	--	已同意	正常	废止

A red arrow points from the '正常' status in the second row to the text above it: '4. 作为接收方，点击授权互通后，一段时间后，通信状态会刷新为“正常”'.

图15 作为接收方，对某条链路授权互通后的状态

对于已批复通过的链路授权可以通过点击“废止”按钮进行取消授权，废止后会导致该条链路中断，网组及后续项目任务无法正常运行。

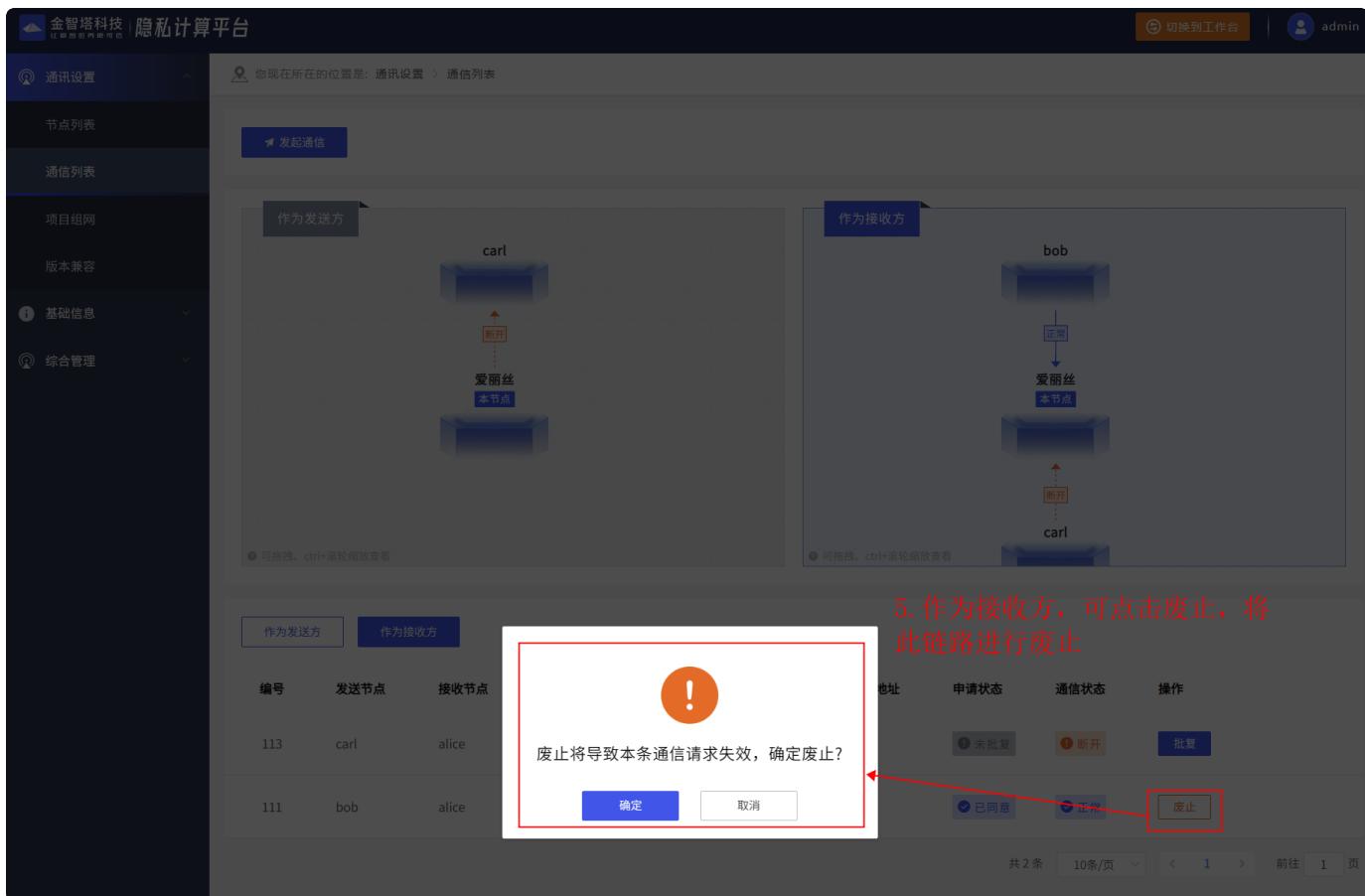


图16 作为接收方，废止链路



6. 作为接收方，废止链路后，其状态更新为：
• 申请状态：废止
• 通信状态：断开（需等待2min）

图17 作为接收方，链路废止后的状态

3.1.3. 项目组网

项目组网模块支持以组长身份拉起网组，网组成员节点必须已经与组长节点完成通信。组网过程中，需要每个组员节点同意加入网组，待网组内的所有组员节点同意后，组长有权利锁定网组。锁定状态下，所有成员不可退出。

点击创建网组，通过拖拽的方式，可以邀请已授权节点列表（项目组网底部列表）中的节点加入网组。

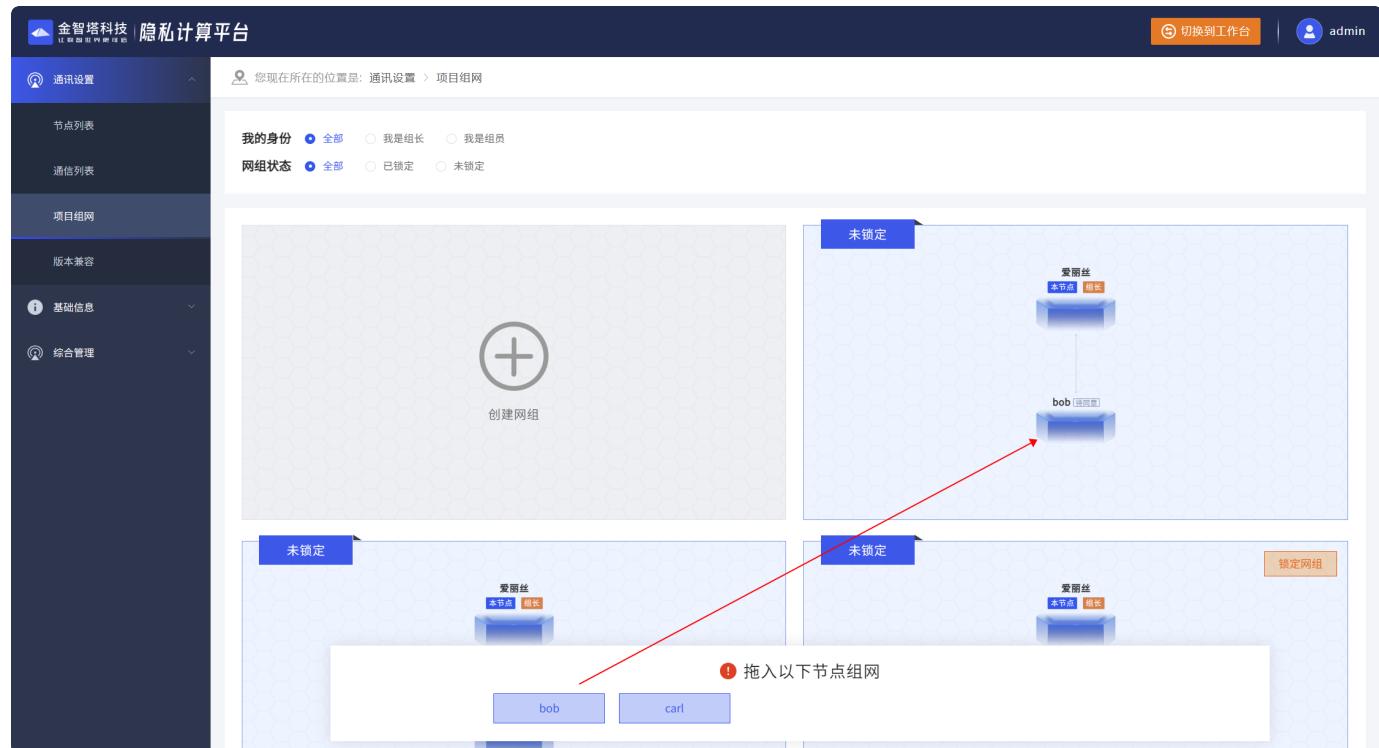


图18 项目组网

邀请之后，被邀请节点的组网界面会显示刚刚创建的网组，可以选择同意或者拒绝。

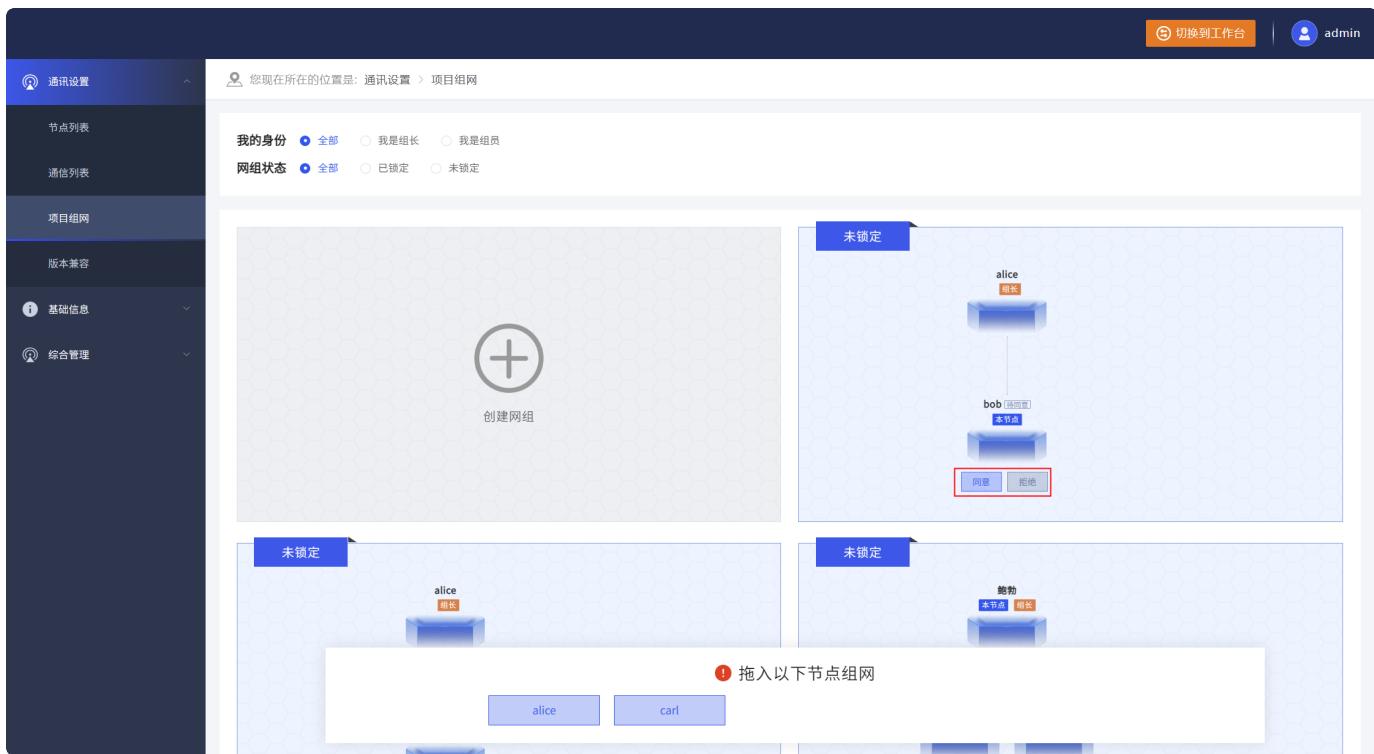


图19 同意组网

当网组内的所有节点都同意之后，网组组长可以锁定网组。锁定状态下，所有成员不可退出。

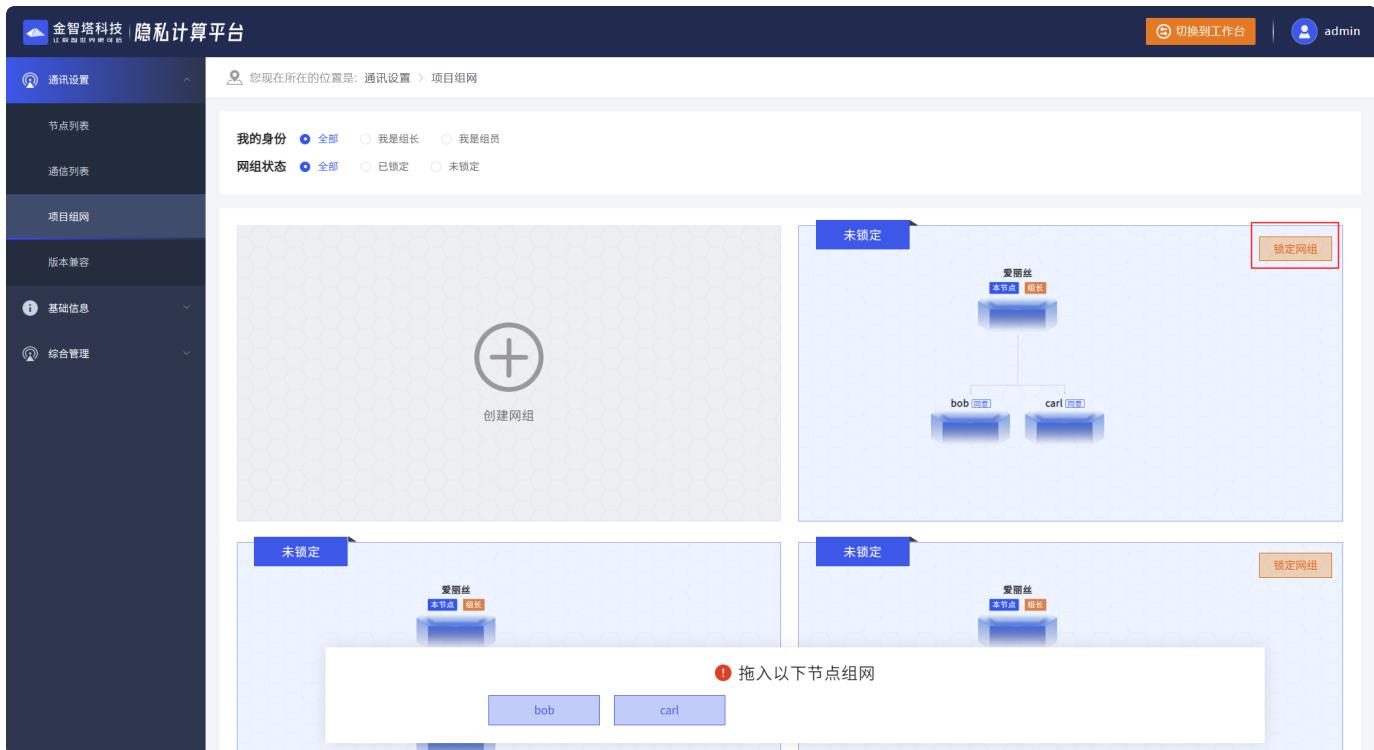


图20 锁定网组

3.1.4. 版本兼容

在版本兼容界面，网组组长可以看到网组下所有节点的版本信息，在创建项目时会自动选择公共最高可用版本。

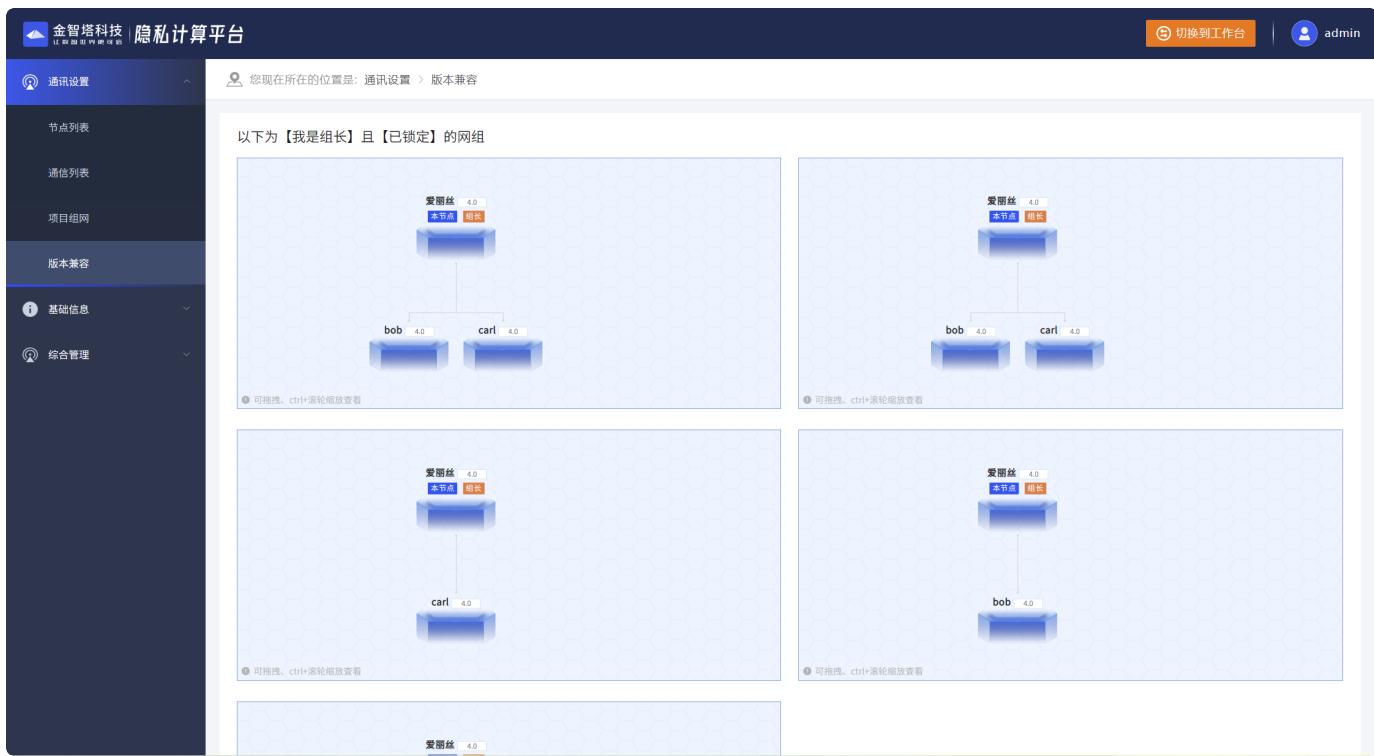


图21 版本兼容

3.2. 基础信息

3.2.1. 管理员账号

管理员账号页面支持修改管理员的用户名名称和账户密码。

* 账号	<input type="text" value="admin"/>
* 用户名称	<input type="text" value="admin"/>
* 密码	<input type="password"/>
* 再次输入	<input type="password"/>

保存账号

图22 管理员账号页面

3.2.2. 节点信息

节点信息页面主要功能：1、导出节点证书，2、配置登入页面的背景图、效果图、标题图、内页logo、license颜色、登入按钮颜色。



图23 登入页图片布局

金智塔科技 隐私计算平台 ← 页内 logo

您现在所在的位置是: 基础信息 > 节点信息

1

2

3

图24 节点信息图

点击上图中一号框中的导出节点证书的按钮，可以导出本机构的节点相关信息，其中包含节点名称，节点唯标识以及节点的公钥信息。点击上图中二号框中的替换按钮，通过上传.jpg类型的图片替换登入页的图片。背景图为登入页的整个背景图，效果图为登入页右侧的展示图，标题图为登入页的右侧的标题图。页内logo为整个平台右上角的logo图。点击上图中三号框，可以修改登入页面的登入按钮和右上角的licence按钮的颜色，避免和登入页背景色颜色混淆。

3.3. 综合管理

3.3.1. 组织设置

用户可以通过管理台-综合管理-组织设置，进行新增组织、在组织下新建角色等操作。

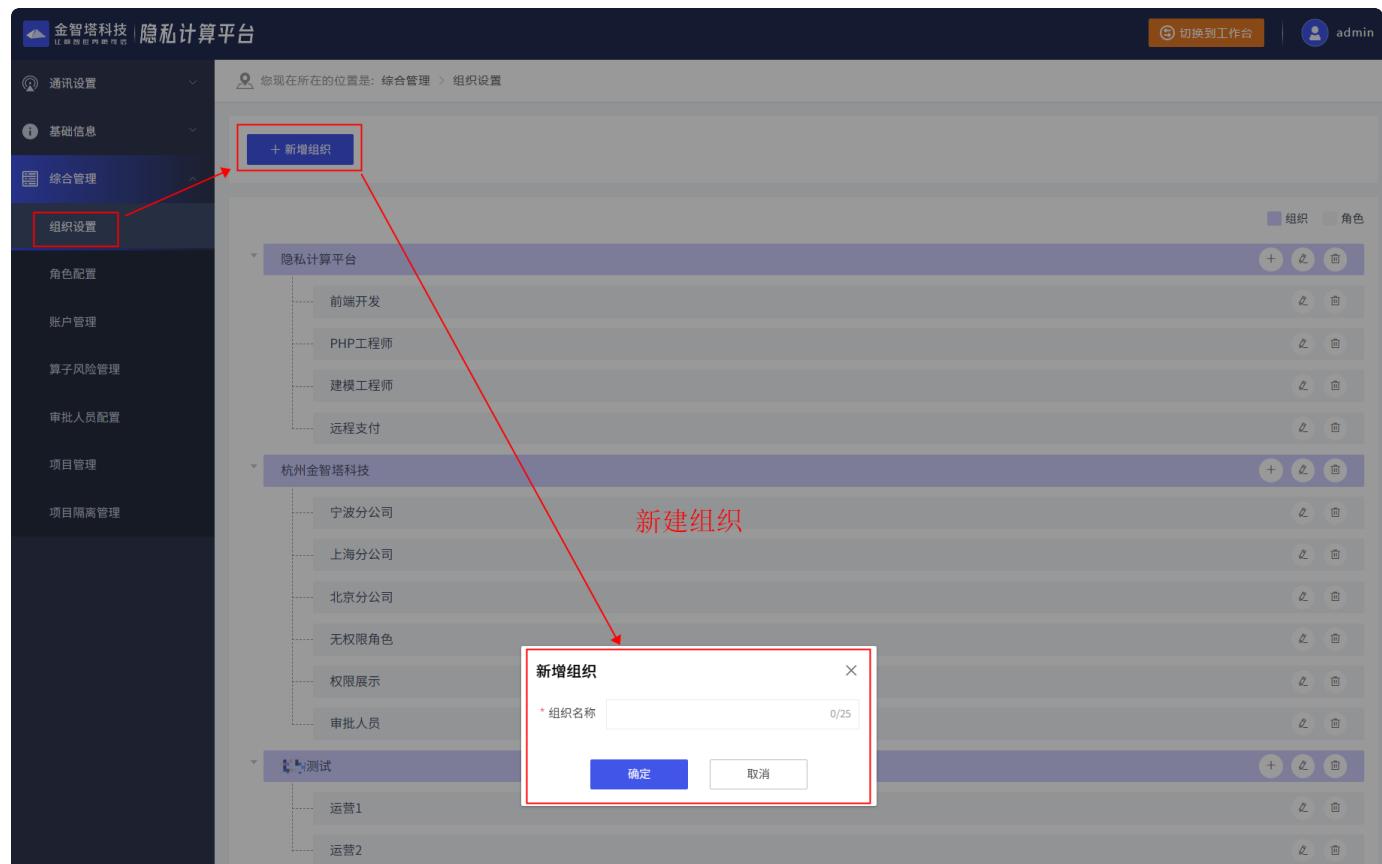


图25 新增组织

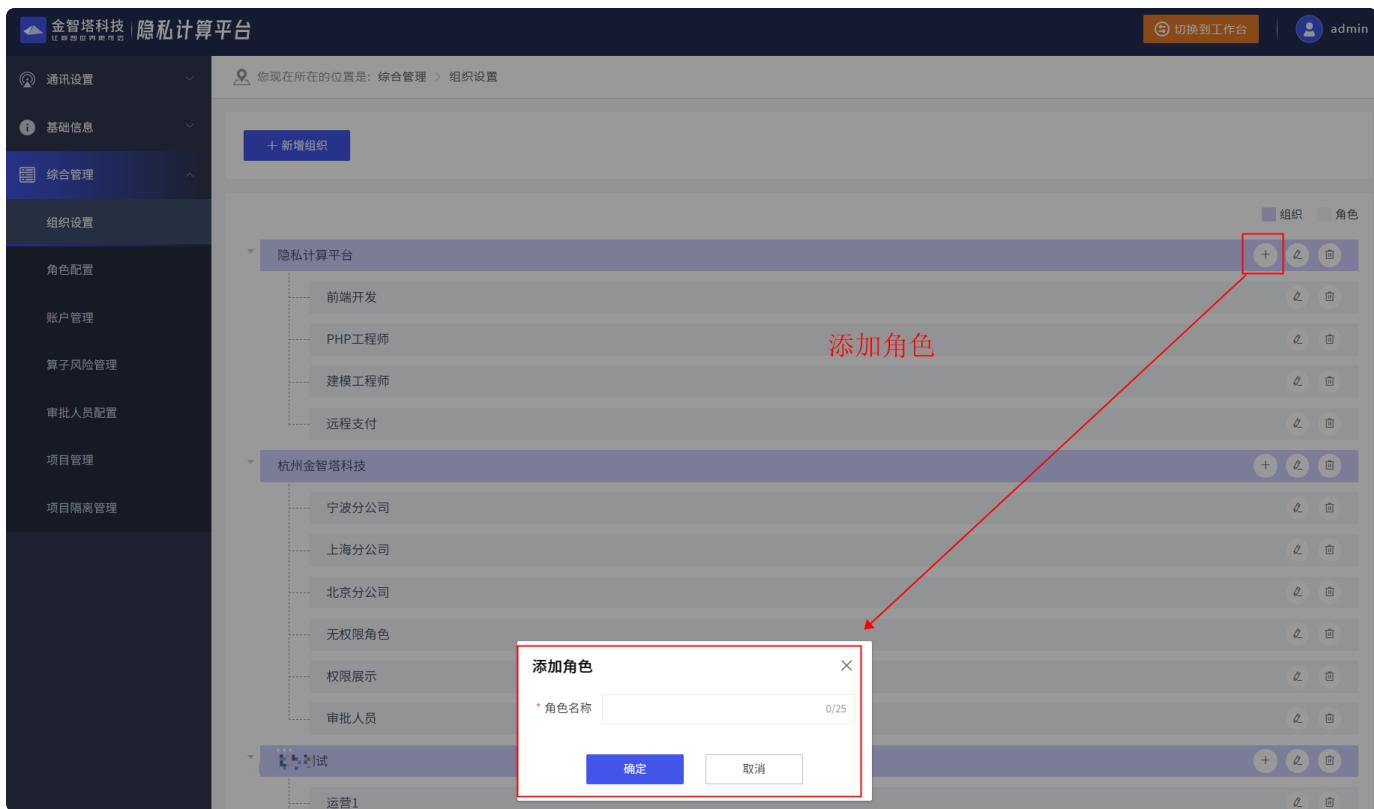


图26 添加角色

3.3.2. 角色配置

用户可对指定组织的角色进行权限的分配，权限的范围包含管理台和工作台的功能键。

序号	角色名称	所属组织	角色权限	状态	操作
91	test001	test2	-	<input checked="" type="checkbox"/> 正常	<button>配置角色</button> <button>冻结</button>
90	PHP	Test	<input type="checkbox"/> 审批人员配置 <input type="checkbox"/> 项目管理 <input type="checkbox"/> 项目隔离管理 <input type="checkbox"/> 任务审核管理 <input type="checkbox"/> 项目管理(工作台) <input type="checkbox"/> 任务管理 <input type="checkbox"/> 数据权限 <input type="checkbox"/> 模型应用	<input checked="" type="checkbox"/> 正常	<button>配置角色</button> <button>冻结</button>
89	java	Test	-	<input checked="" type="checkbox"/> 正常	<button>配置角色</button> <button>冻结</button>
88	审批人员	杭州金智塔科技	<input type="checkbox"/> 任务审核管理	<input checked="" type="checkbox"/> 正常	<button>配置角色</button> <button>冻结</button>
86	远程支付	隐私计算平台	<input type="checkbox"/> 项目管理(工作台) <input type="checkbox"/> 数据权限	<input checked="" type="checkbox"/> 正常	<button>配置角色</button> <button>冻结</button>
85	权限展示	杭州金智塔科技	<input type="checkbox"/> 审批人员配置 <input type="checkbox"/> 项目管理 <input type="checkbox"/> 项目隔离管理 <input type="checkbox"/> 任务审核管理 <input type="checkbox"/> 项目管理(工作台) <input type="checkbox"/> 任务管理 <input type="checkbox"/> 数据权限 <input type="checkbox"/> 模型应用	<input checked="" type="checkbox"/> 正常	<button>配置角色</button> <button>冻结</button>
83	无权限角色	杭州金智塔科技	<input type="checkbox"/> 项目管理	<input checked="" type="checkbox"/> 正常	<button>配置角色</button> <button>冻结</button>
80	运营2	杭州金智塔科技	-	<input checked="" type="checkbox"/> 正常	<button>配置角色</button> <button>冻结</button>
79	运营1	杭州金智塔科技	-	<input checked="" type="checkbox"/> 正常	<button>配置角色</button> <button>冻结</button>
78	北京分公司	杭州金智塔科技	-	<input checked="" type="checkbox"/> 正常	<button>配置角色</button> <button>冻结</button>

共 15 条 | 10条/页 | < 1 2 > 前往 1 页

图27 角色配置

3.3.3. 账户管理

在综合管理-账户创建页面可以支持当前节点的账户新增、删除、修改、密码重置及冻结解冻等功能。

金智塔科技 隐私计算平台

切换到工作台 | admin

通讯设置 基础信息 综合管理 账户管理 算子风险管理 审批人员配置 项目管理 项目隔离管理

您现在所在的位置是: 综合管理 > 账户管理

1. 创建账户

2. 对创建后的账户进行修改等操作

序号	用户名	账号	组织/角色	状态	操作
60			Test/PHP	正常	<button>修改</button> <button>删除</button> <button>...</button>
59	二级审批人员4	sp4	杭州金智塔科技/审批人员	正常	<button>修改</button> <button>删除</button> <button>重置密码</button>
58	二级审批人员3	sp3	杭州金智塔科技/审批人员	正常	<button>修改</button> <button>删除</button> <button>冻结</button>
57	一级审批人员2	sp2	杭州金智塔科技/审批人员	正常	<button>修改</button> <button>删除</button> <button>...</button>
56	一级审批人员1	sp1	杭州金智塔科技/审批人员	冻结	<button>修改</button> <button>删除</button> <button>...</button>
55	111	111	杭州金智塔科技/宁波分公司	正常	<button>修改</button> <button>删除</button> <button>...</button>
54	12	12	杭州金智塔科技/北京分公司	正常	<button>修改</button> <button>删除</button> <button>...</button>
53	11	11	杭州金智塔科技/北京分公司	正常	<button>修改</button> <button>删除</button> <button>...</button>
52	10	10	杭州金智塔科技/北京分公司	正常	<button>修改</button> <button>删除</button> <button>...</button>
51	9	9	杭州金智塔科技/北京分公司	正常	<button>修改</button> <button>删除</button> <button>...</button>

共 19 条 10条/页 < 1 2 > 前往 1 页

图28 账户列表

金智塔科技 隐私计算平台

切换到工作台 | admin

通讯设置 基础信息 综合管理 账户管理 算子风险管理 审批人员配置 项目管理 项目隔离管理

您现在所在的位置是: 综合管理 > 账户管理

1. 创建账户时，可设置用户名和账密

2. 创建账户时，可配置组织和角色类型

创建账户

* 用户姓名: chengrj
* 账号: chengrj
* 密码设置:
* 确认密码:
* 所属组织: 隐私计算平台
* 角色设置: 前端开发 (checked)
建模工程师
PHP 工程师
远程支付

确定 取消

图29 账户创建

3.3.4. 算子风险管理

用户可在算子风险管理，对算子进行风险设置，可将低风险算子转移为高风险算子，也可将高风险算子转移为低风险算子。

金智塔科技 隐私计算平台

您现在所在的位置是：综合管理 > 算子风险管理 *高风险算子被使用时，需要审核人员审核画布任务

恢复默认设置

一级分类	二级分类	低风险算子（被使用免审核）	高风险算子（被使用需审核）
数据处理	预处理	字符串转化 稀疏样本删除 缺失特征删除 缺失值填充	
	特征工程	特征分箱 卡方分箱 类别编码 one-hot编码	
	自定义加工	SQL编辑 Python编辑	点击转移 ↔
数据分析		皮尔逊相关系数 信息值IV 特征重要性分析 方差膨胀系数VIF	
			联邦求交(FL) 多方安全求交(DH) 联邦求差
联合建模	联邦学习(FL)	神经网络 逻辑回归 随机森林 决策树 Xgboost K-means	
		多方安全计算(MPC)	逻辑回归(MPC) Xgboost(MPC) 线性回归
			安全加法 安全减法 安全乘法 安全除法 安全比较
基础运算			最大值 最小值 中位数 均值 方差
			匿名查询(PIR) 加密传输
统计运算			偏离度计算 自定义指标
隐私查询			
特殊运算			
业务算法		评分卡	
样本筛选		保留样本 剔除样本	

图30 低风险算子转移至高风险算子

金智塔科技 隐私计算平台

您现在所在的位置是：综合管理 > 算子风险管理 *高风险算子被使用时，需要审核人员审核画布任务

恢复默认设置

一级分类	二级分类	低风险算子（被使用免审核）	高风险算子（被使用需审核）
数据处理	预处理	字符串转化 稀疏样本删除 缺失特征删除 缺失值填充	
	特征工程	特征分箱 卡方分箱 类别编码 标准化	
	自定义加工	SQL编辑 Python编辑	
数据分析		皮尔逊相关系数 信息值IV 特征重要性分析 方差膨胀系数VIF	
			多方安全求交(DH) 联邦求差
联合建模	联邦学习(FL)	神经网络 逻辑回归 随机森林 决策树 Xgboost K-means	
		多方安全计算(MPC)	逻辑回归(MPC) Xgboost(MPC) 线性回归
			点击转移 ↔
基础运算			安全加法 安全减法 安全乘法 安全除法 安全比较
统计运算			最大值 最小值 中位数 均值 方差
隐私查询			匿名查询(PIR) 加密传输
特殊运算			偏离度计算 自定义指标
业务算法		评分卡	
样本筛选		保留样本 剔除样本	

图31 高风险算子转移至低风险算子

3.3.5. 审批人员配置

用户可在审批人员配置页面，设置审评流程，其中流程链路支持多级审批，同时每一级审批支持多人审批。



图32 审批流程配置

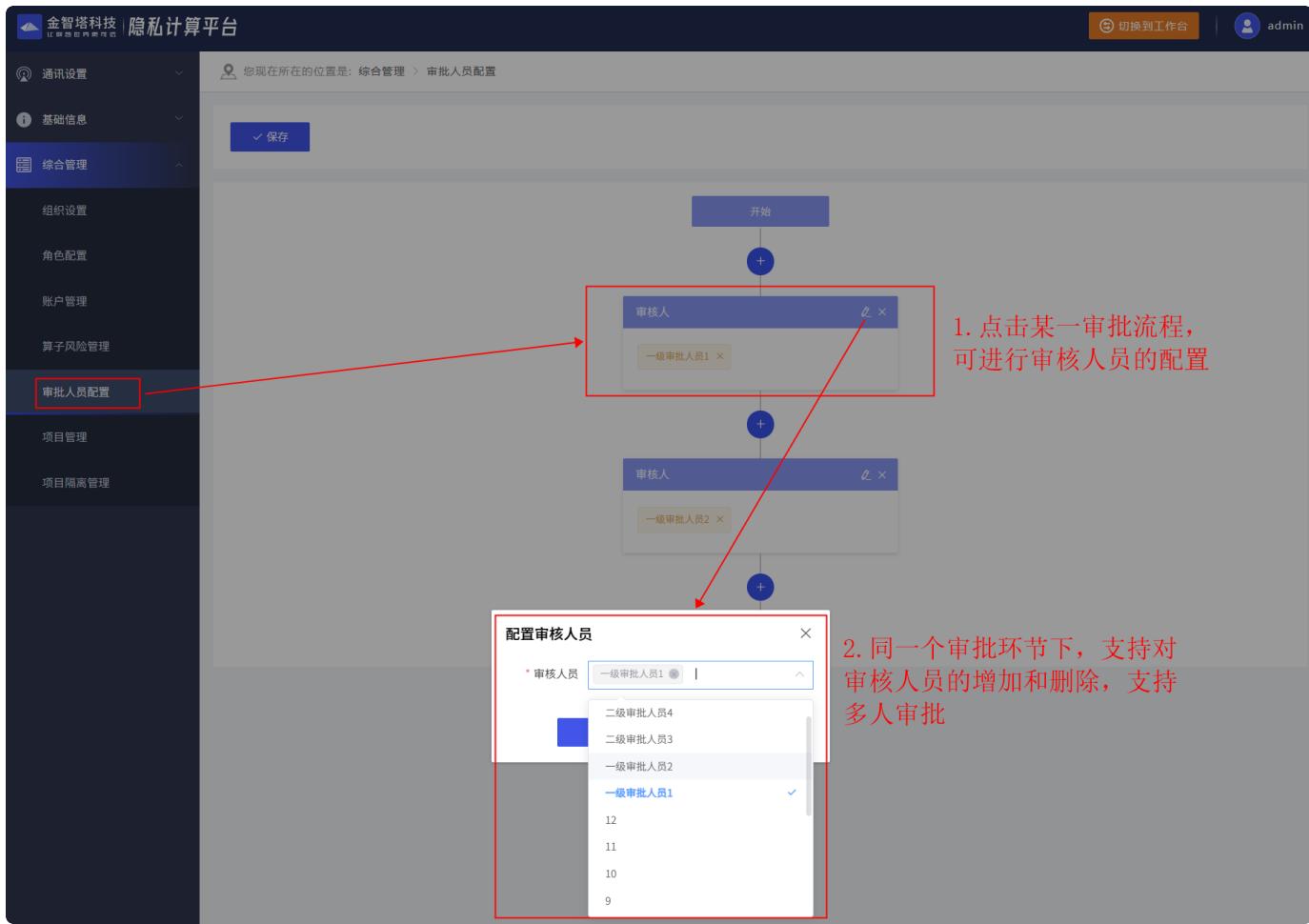


图33 审批人员配置

3.3.6. 项目管理

3.3.6.1. 合作方项目列表

作为数据提供方，可以给各项目提供数据、审批数据权限。

编号	项目名称	参与身份	需求单位	创建时间	操作
PbobM8GUH	wangzy/测试项目	数据提供方	鲍勃	2023-01-18 17:09:04	<button>数据外供审批</button> <button>所属网组</button>
PbobSfFeP	bobp001	数据提供方	鲍勃	2023-01-18 14:28:10	<button>数据外供审批</button> <button>所属网组</button>
PbobnQHol	Test	数据提供方	鲍勃	2023-01-17 14:29:21	<button>数据外供审批</button> <button>所属网组</button>
PcarlY3IVg	正常数据集	数据提供方	carl	2023-01-12 15:17:50	<button>数据外供审批</button> <button>所属网组</button>
Pbob1OjvR	测试	数据提供方	bob	2023-01-12 14:44:44	<button>数据外供审批</button> <button>所属网组</button>
Pcarlmyuhn	carltoalice	数据提供方	carl	2023-01-09 14:00:20	<button>数据外供审批</button> <button>所属网组</button>
PbobcLzLp	zkm-test	数据提供方	bob	2023-01-09 10:31:21	<button>数据外供审批</button> <button>所属网组</button>
Pbobfd3Bc	普通用户创建的项目	数据提供方	bob	2023-01-09 10:25:11	<button>数据外供审批</button> <button>所属网组</button>
PbobideS7	测试相关管理创建人员	数据提供方	bob	2023-01-09 10:23:23	<button>数据外供审批</button> <button>所属网组</button>
Pbobsyppn	test普通用户创建	数据提供方	bob	2023-01-07 16:08:26	<button>数据外供审批</button> <button>所属网组</button>

共 13 条 10条/页 < 1 2 > 前往 1 页

图34 项目列表

用户可以点击“所属网组”，查看该网组成员。

编号	项目名称	参与身份	需求单位	创建时间	操作
PbobM8GUH				2023-01-18 17:09:04	<button>数据外供审批</button> <button>所属网组</button>
PbobSfFeP				2023-01-18 14:28:10	<button>数据外供审批</button> <button>所属网组</button>
PbobnQHol				2023-01-17 14:29:21	<button>数据外供审批</button> <button>所属网组</button>
PcarlY3IVg				2023-01-12 15:17:50	<button>数据外供审批</button> <button>所属网组</button>
Pbob1OjvR				2023-01-12 14:44:44	<button>数据外供审批</button> <button>所属网组</button>
Pcarlmyuhn				2023-01-09 14:00:20	<button>数据外供审批</button> <button>所属网组</button>
PbobcLzLp				2023-01-09 10:31:21	<button>数据外供审批</button> <button>所属网组</button>
Pbobfd3Bc				2023-01-09 10:25:11	<button>数据外供审批</button> <button>所属网组</button>
PbobideS7				2023-01-09 10:23:23	<button>数据外供审批</button> <button>所属网组</button>
Pbobsyppn				2023-01-07 16:08:26	<button>数据外供审批</button> <button>所属网组</button>

共 13 条 10条/页 < 1 2 > 前往 1 页

图35 所属网组

用户可以在**数据外供审批**页面中，查看项目的基本信息，例如项目名称、项目简介、所属网组、组长单位、创建单位、创建人员、创建时间等，还能查看该项目可使用的组件。

The screenshot shows the 'Wangzy Test Project' page. At the top, there's a header with the project name and a red box highlighting the main content area. Below the header, there's a table with project details:所属网组 (NaliceQc3VF), 组长单位 (爱丽丝), 创建人 (鲍勃), 创建时间 (2023-01-18 17:09:04). The main content area is divided into sections: '项目算子' (Project Operators) with categories like '一级分类' (Primary Category), '二级分类' (Secondary Category), and '算子列表' (Operator List); '基本信息' (Basic Information) showing file details (89727307511246_188.csv, 10000 rows, CSV type); and '安全设置' (Security Settings) with a status table.

图36 项目基本信息

3.3.6.2. 数据对外提供

用户可以给某个项目提供数据集，有新增表和新增CSV两种方式。

新增表包括Mysql、PostgreSQL、DB2、Oracle、SQL Server等方式，通过输入数据库信息可以查询到表，选择对应表即可进行库表登录信息的导入，如下图所示。

The screenshot shows the 'Add New Table' dialog box. It has fields for '数据源名称' (Data Source Name), '数据库类型' (Database Type), '数据库地址' (Database Address), '用户名' (Username), '密码' (Password), '数据库名' (Database Name), 'schema' (Schema), and '表名' (Table Name). A red arrow points from the '新增表' (Add New Table) button on the left to the '数据库类型' dropdown menu, which is expanded to show options: Mysql, PostgreSQL, DB2, Oracle, and SQL Server. The background shows the project's basic information and security settings.

图37 新增数据表

新增CSV的模式下，输入数据源名称后，选择本地CSV文件进行上传。上传成功后，会显示文件在节点所部署的服务器中的实际存储路径，点击新增，即可完成CSV文件的导入，如下图所示。

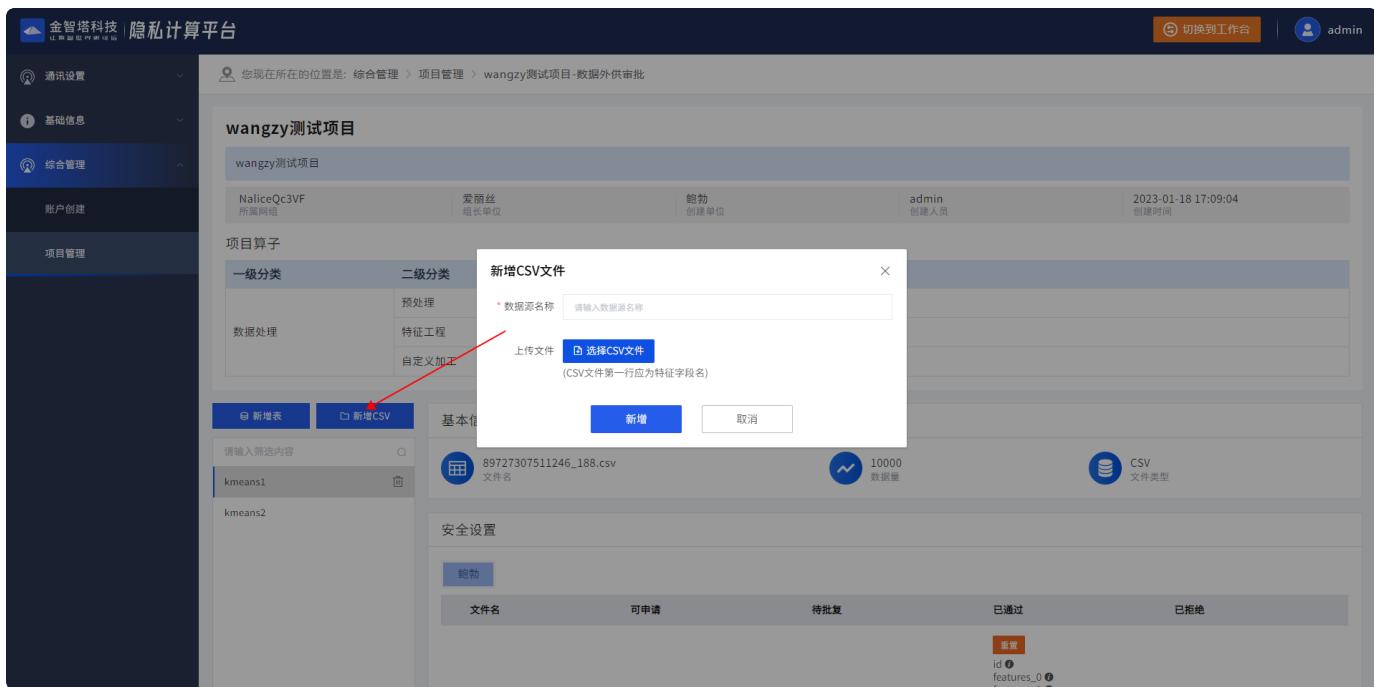


图38 新增CSV

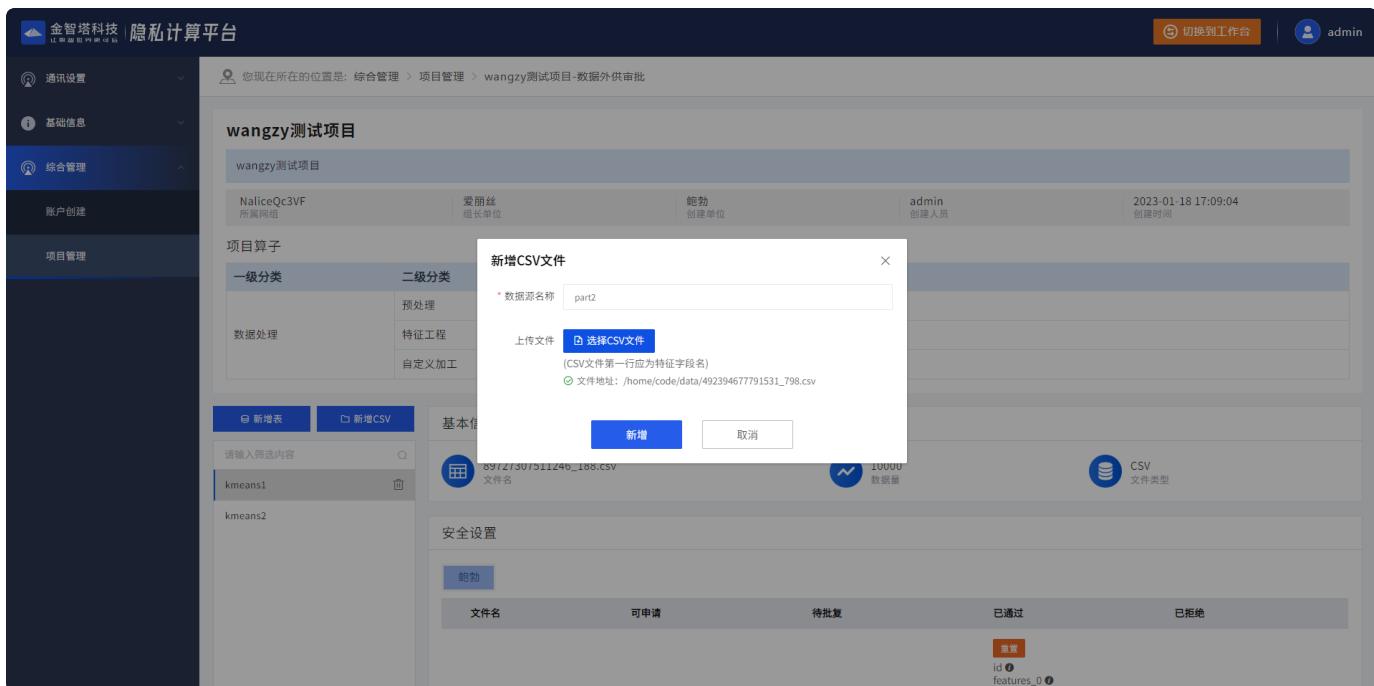


图39 CSV文件上传成功

用户可以在**数据外供审批**页面中，点击指定数据集后，进行数据字典的查看及修改完善内容，包括数据中文名、数据类型、数据枚举、注释等内容的完善。

您现在所在的位置是: 综合管理 > 项目管理 > wangzy测试项目-数据外供审批

数据字典

数据项名	中文名	数据类型	注释	取值范围/类别全枚举	操作
id	ID	int			完善数据字典
features_0	特征0	varchar			完善数据字典
features_1	特征1	varchar			完善数据字典
features_2	特征2	varchar			完善数据字典
features_3	特征3	varchar			完善数据字典

共 252 条 5条/页 < 1 2 3 4 5 6 ... 51 > 前往 1 页

图40 数据字典详情

您现在所在的位置是: 综合管理 > 项目管理 > wangzy测试项目-数据外供审批

完善数据字典

数据项名	中文名	数据类型	注释	取值范围/类别全枚举	操作
id					完善数据字典
features_0	特征0	varchar			完善数据字典
features_1	特征1	varchar			完善数据字典
features_2	特征2	varchar			完善数据字典
features_3	特征3	varchar			完善数据字典

共 252 条 5条/页 < 1 2 3 4 5 6 ... 51 > 前往 1 页

图41 完善数据字典

用户可以在**数据外供审批**页面中，查看自己已经导入成功的数据目录，点击任意一个数据集都能查看对应的基本信息，安全设置及完善数据字典等操作。数据列表支持通过输入筛选内容，筛选出对应的数据集。

The screenshot shows the 'Data Directory Management' section of the platform. At the top, there's a search bar with placeholder text '您现在所在的位置是: 综合管理 > 项目管理 > wangzy测试项目-数据外供审批'. Below the search bar is a title 'wangzy测试项目'. A table displays project details:所属网组 'NaliceQc3VF', 组长单位 '爱丽丝', 责任 '鲍勃', 创建人 'admin', 创建时间 '2023-01-18 17:09:04'. The main area is titled '项目算子' and contains a grid of operators categorized by type (一级分类) and sub-type (二级分类). The '算子列表' section includes buttons for '新增表' and '新增CSV'. A red box highlights the search input field and dropdown menu.

图42 数据目录管理

3.3.6.3. 数据审批

用户可以在**数据外供审批**页面中，点击具体数据源时，会展示合作方节点对该数据的申请情况，并进行数据使用权限的审批，也可以将审批通过/拒绝的字段进行重置。

The screenshot shows the 'Data Audit Operation' section. It features a search bar with placeholder text '您现在所在的位置是: 综合管理 > 项目管理 > wangzy测试项目-数据外供审批'. Below the search bar is a title 'wangzy测试项目'. A table displays basic information about a data source: 文件名 '89727307511246_188.csv', 数据量 '10000', and 文件类型 'CSV'. The main area is titled '安全设置' and contains a table of access requests. A red box highlights the '通过' (Approve) and '拒绝' (Reject) buttons for a specific row. Another red box highlights the '重置' (Reset) button at the bottom right of the table.

图43 数据审批操作

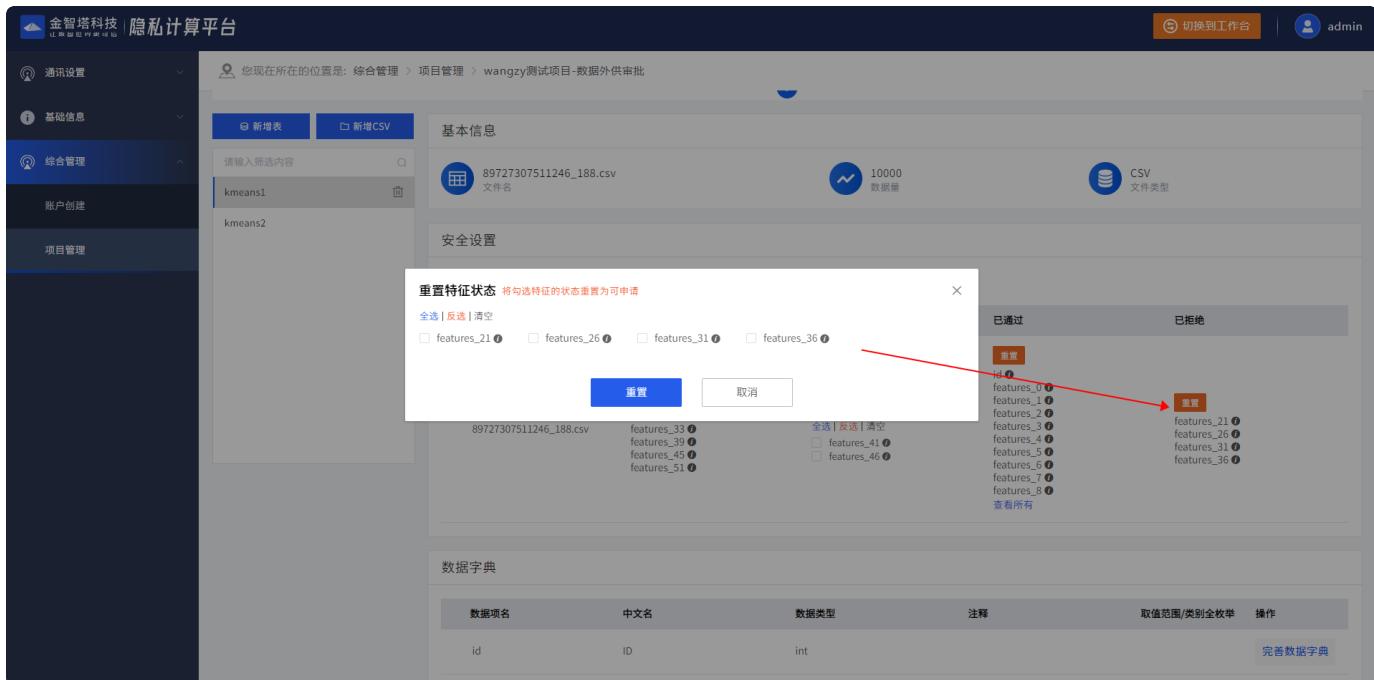


图44 重置状态

3.3.7. 项目隔离管理

项目隔离管理支持三种模型：项目仅对创建人员开放、项目仅对同组织成员开放、项目仅同节点成员开放。



图45 项目隔离管理

3.3.8. 任务审核管理

当3.3.5. 审批人员配置中对某账号进行了审批流程的配置，则该账号的管理台的综合管理下，会出现任务审核管理。该功能可对高风险算子的申请下载数据操作进行审核。

The screenshot shows the 'Task Audit Management' section of the platform. At the top, there's a header with the logo '金智塔科技' and the text '隐私计算平台'. On the right, it says '二级审批人员3'. Below the header, a breadcrumb navigation shows '综合管理 > 任务审核管理 *对含高风险算子的任务审核'. A left sidebar has a '任务审核管理' tab selected. The main content area has columns for '任务编号', '任务名称', '需求单位', '审核状态', '审核说明', '审核时间', and '操作'. In the center, there's a small icon of a briefcase with an arrow pointing up and the text '暂无数据'. At the bottom, there are pagination controls showing '共0条' and '10条/页'.

图46 任务审核管理

4.工作台

隐私计算工作台负责具体的项目执行工作，建立一个项目需要完成数据权限申请、项目编辑、模型应用等流程中的一个或几个。在功能上可满足用户完成数据处理、数据分析、基础运算、统计运算、特殊运算、隐私查询、联合建模、联合SQL、模型发布管理等工作。

4.1. 项目管理

4.1.1 创建项目

用户可在平台右上方点击创建项目来新建一个项目，然后可对项目名称、项目描述进行编辑，以及选择网组（只有锁定的网组才能创建项目）、选择算子。点击确定后，可通过点击项目编辑来进入该项目。



图47 创建项目

The screenshot shows the Jinzhitech Privacy Computing Platform interface. At the top, there are two project cards: 'p002' and 'caicai'. Below them is a modal window titled '选择算子' (Select Operator). The modal has three tabs: '全选' (Select All), '清空' (Clear), and '算子列表' (Operator List). The '算子列表' tab is active and displays a grid of operators categorized by primary and secondary type. A red box highlights the '算子列表' section. At the bottom of the modal are '确定' (Confirm) and '取消' (Cancel) buttons.

图48 选择算子

4.1.2 创建任务

用户可以在创建好的项目中（进入项目编辑），通过点击创建任务来新建一个项目，在对任务名称进行编辑后，点击确定即可新增一个任务，然后点击画布编辑即可进入该任务的工作台页面，用户可进行具体任务的拖拉拽操作。

金智塔科技 | 隐私计算平台

切换到管理台 | admin

创建项目

p002

bob carl 1111

鲍勃 创建单位 admin 创建人员 2023-01-14 15:27:34 创建时间

NbobVzsS1 所属网组 > 鲍勃 组长单位

数据权限 项目编辑 模型应用

caicai

test

鲍勃 创建单位 admin 创建人员 2023-01-13 10:07:04 创建时间

NaliceFrzK1 所属网组 > 爱丽丝 组长单位

数据权限 项目编辑 模型应用

pwq test

test

鲍勃 创建单位 admin 创建人员 2023-01-13 09:46:42 创建时间

NaliceFrzK1 所属网组 > 爱丽丝 组长单位

数据权限 项目编辑 模型应用

点击项目编辑，进入任务列表

图49 项目编辑

金智塔科技 | 隐私计算平台

切换到管理台 | admin

您现在所在的位置是: 项目列表 > p002

bob carl 1111

请输入任务名称

创建任务

* 任务名称

确定 取消

001-lr

创建者 admin

更新时间 2023-01-17 17:29:33

需求节点 鲍勃

画布编辑

点击创建任务，编辑名称

图50 创建任务



图51 画布编辑

4.1.3 更多操作

用户可以在任务列表页面，选择某个任务，在右上角的更多操作中，**复制任务、编辑任务名称和删除任务**。其中复制任务，会将任务的画布内容进行复制；删除任务时，会进行弹窗确认，确认无误后点击确定按钮进行任务删除。



图52 更多操作

4.2. 数据权限

4.2.1. 数据提供

用户可以在工作台中，点击指定项目的**数据权限**，进行数据权限管理页面，为当前项目提供本方数据。

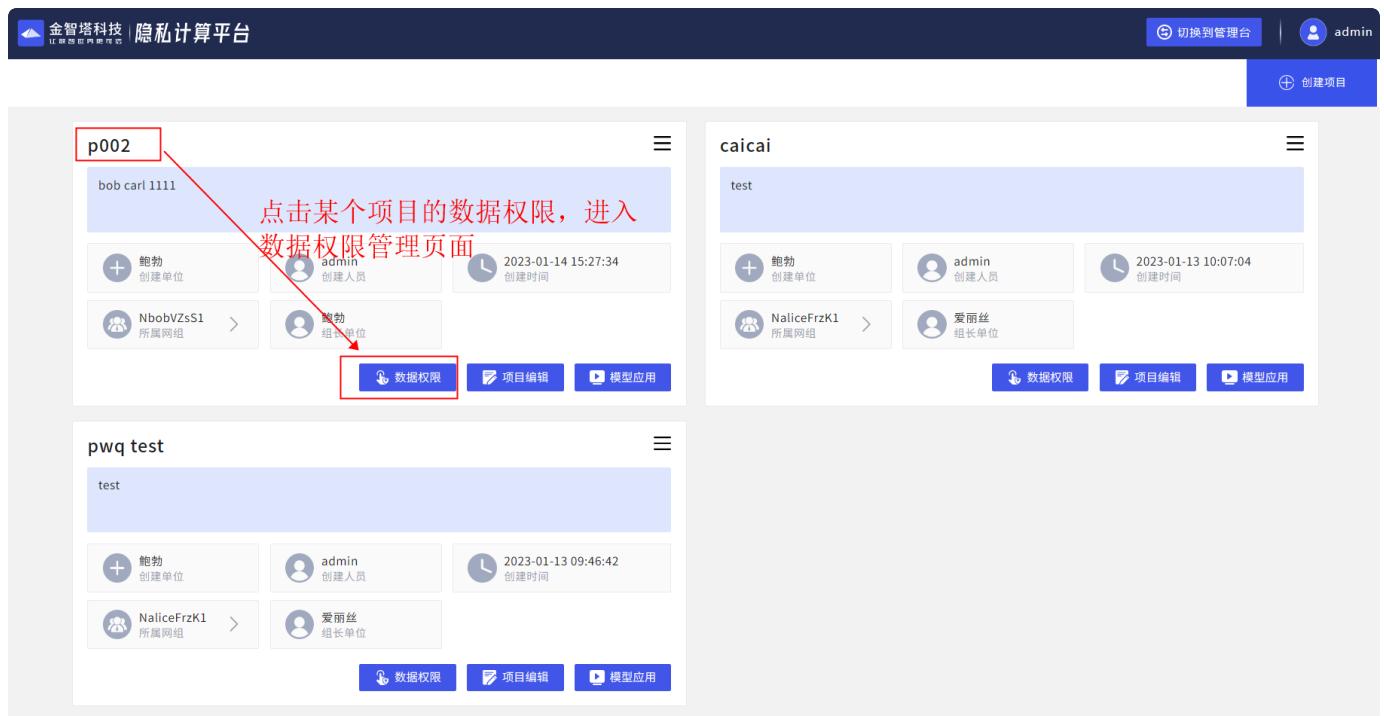


图53 数据权限

提供数据的方式有两种方式：数据库导入和CSV导入。其中数据库导入支持Mysql、PostgreSQL、DB2、Oracle、SQL Server等主流数据库，通过输入数据库信息可以查询到表，选择对应表即可进行库表登录信息的导入，如下图所示。

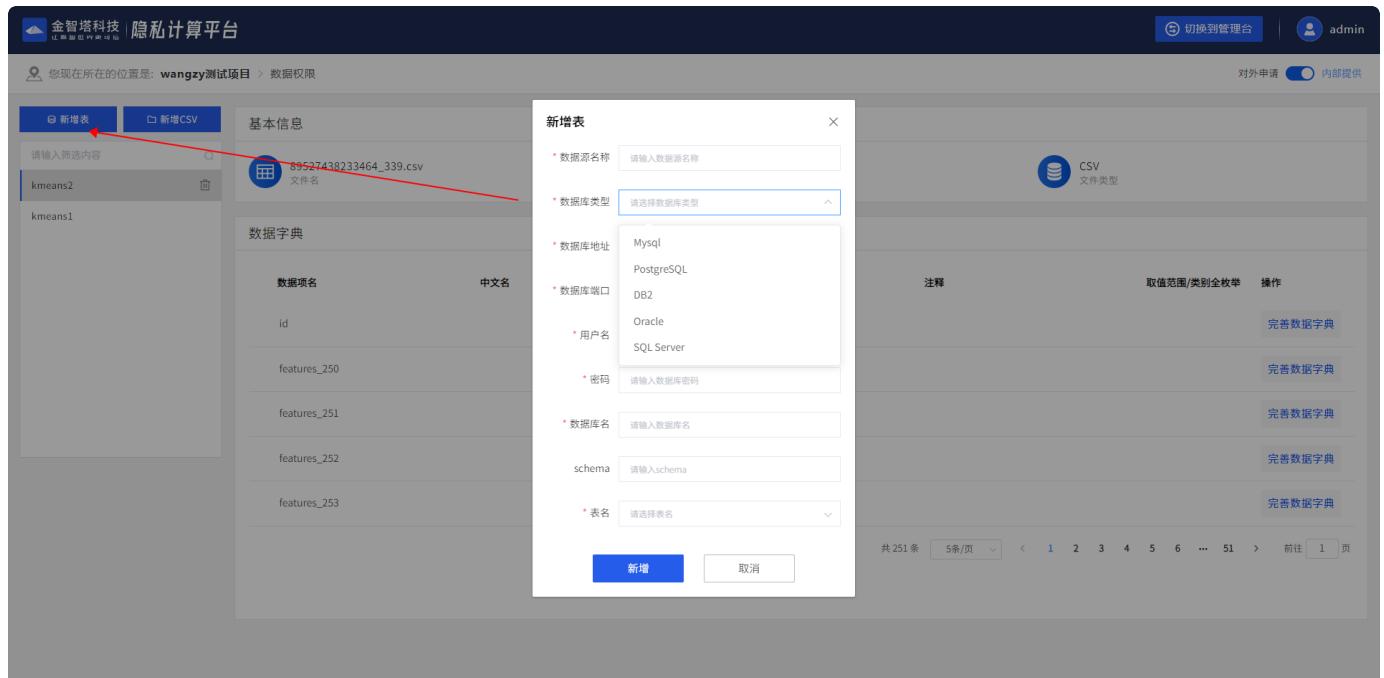


图54 新增数据表

新增CSV的模式下，输入数据源名称后，选择本地CSV文件进行上传。上传成功后，会显示文件在节点所部署的服务器中的实际存储路径，点击新增，即可完成CSV文件的导入，如下图所示。

This screenshot shows the 'Add CSV' interface in the platform. At the top, there are two buttons: '新增表' (New Table) and '新增CSV' (New CSV). The '新增CSV' button is highlighted with a red arrow. Below it, there's a section for 'Basic Information' showing a file named '89527438233464_339.csv' with a size of 10000 bytes and a CSV file type. A modal window titled 'Add CSV File' is open, prompting for a data source name and providing a 'Select CSV File' button. The background shows a data dictionary table with columns like 'Data Item Name', 'Chinese Name', 'Data Type', and 'Annotations'.

图55 新增CSV入口

This screenshot shows the 'Select CSV File' interface. It displays a table of data items with columns for 'Data Item Name', 'Chinese Name', 'Data Type', and 'Annotations'. A modal window titled 'Add CSV File' is overlaid, with step 1 '选择csv文件' (Select CSV file) pointing to the 'Select CSV File' button and step 2 '上传成功后, 可以点击开始上传' (After upload success, click to start upload) pointing to the 'Start Upload' button. The background table has rows for 'ID', 'loanAmnt', 'term', 'interestRate', and 'installment'.

图56 选择csv文件

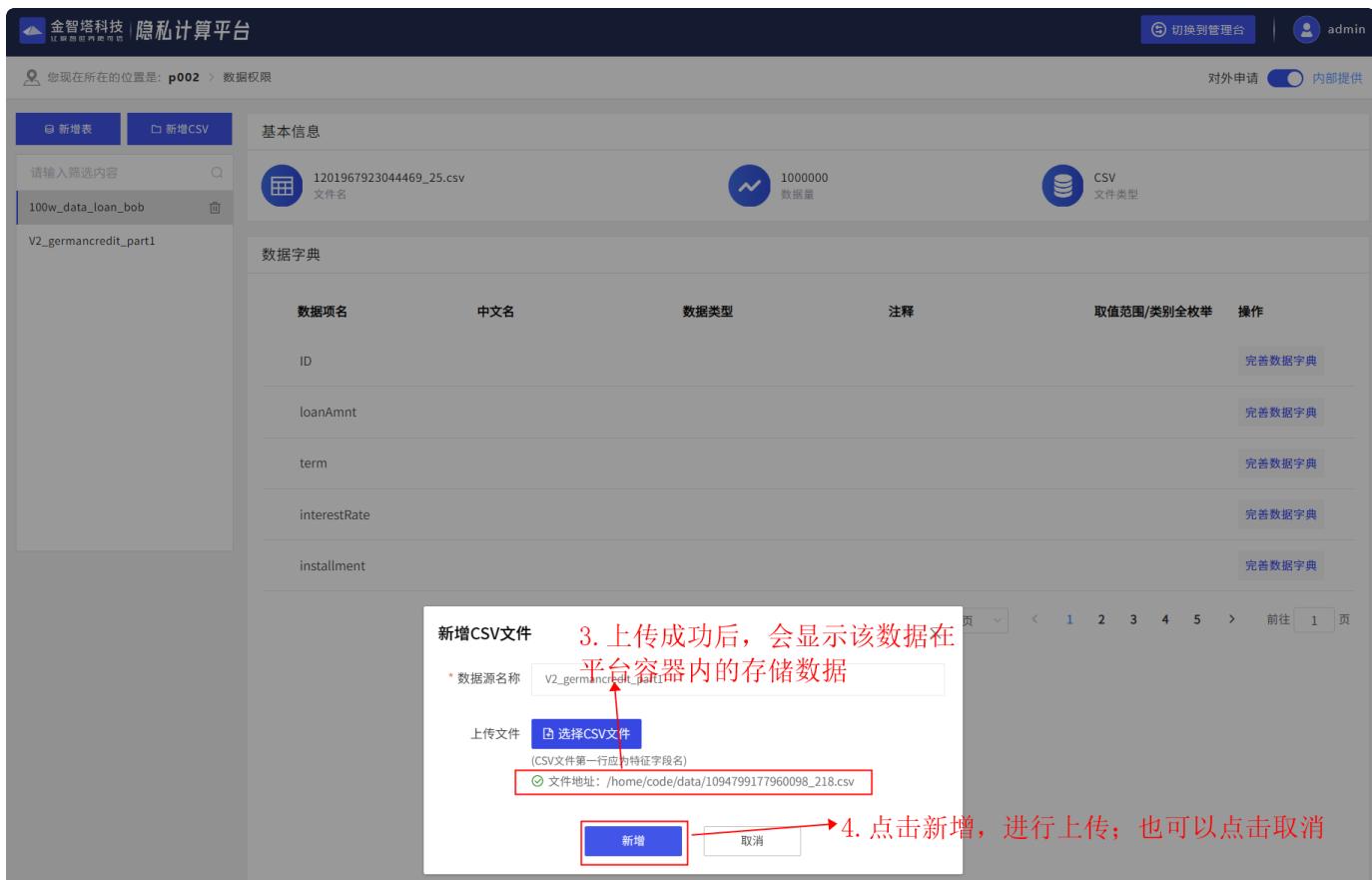


图57 上传成功和点击新增

用户可以在数据权限管理页面中，点击指定数据集后，进行数据字典的查看及修改完善内容，包括数据中文名、数据类型、数据枚举、注释等内容的完善。

数据项名	中文名	数据类型	注释	取值范围/类别全枚举	操作
id					完善数据字典
features_250					完善数据字典
features_251					完善数据字典
features_252					完善数据字典
features_253					完善数据字典

图58 数据字典详情

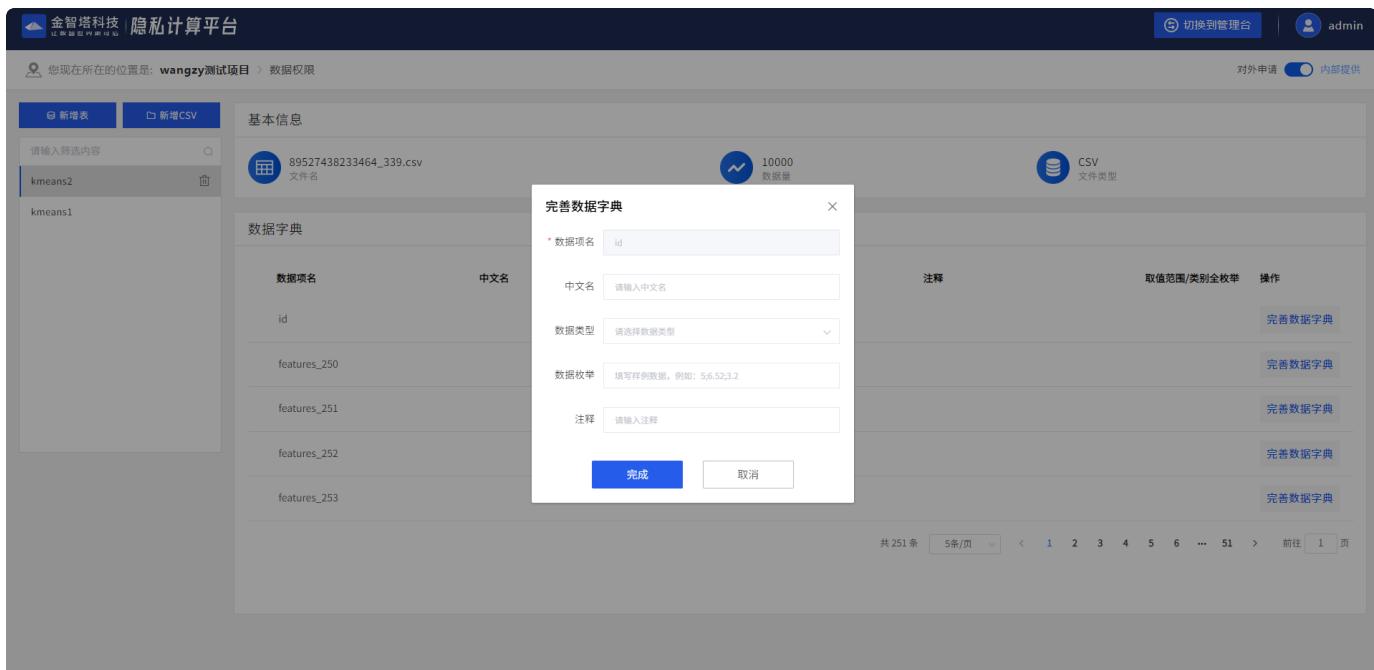


图59 完善数据字典

用户可以在数据权限管理页面中，查看自己已经导入成功的数据目录，点击任意一个数据集都能查看对应的基本信息，完善数据字典等操作。数据列表支持通过输入筛选内容，筛选出对应的数据集。

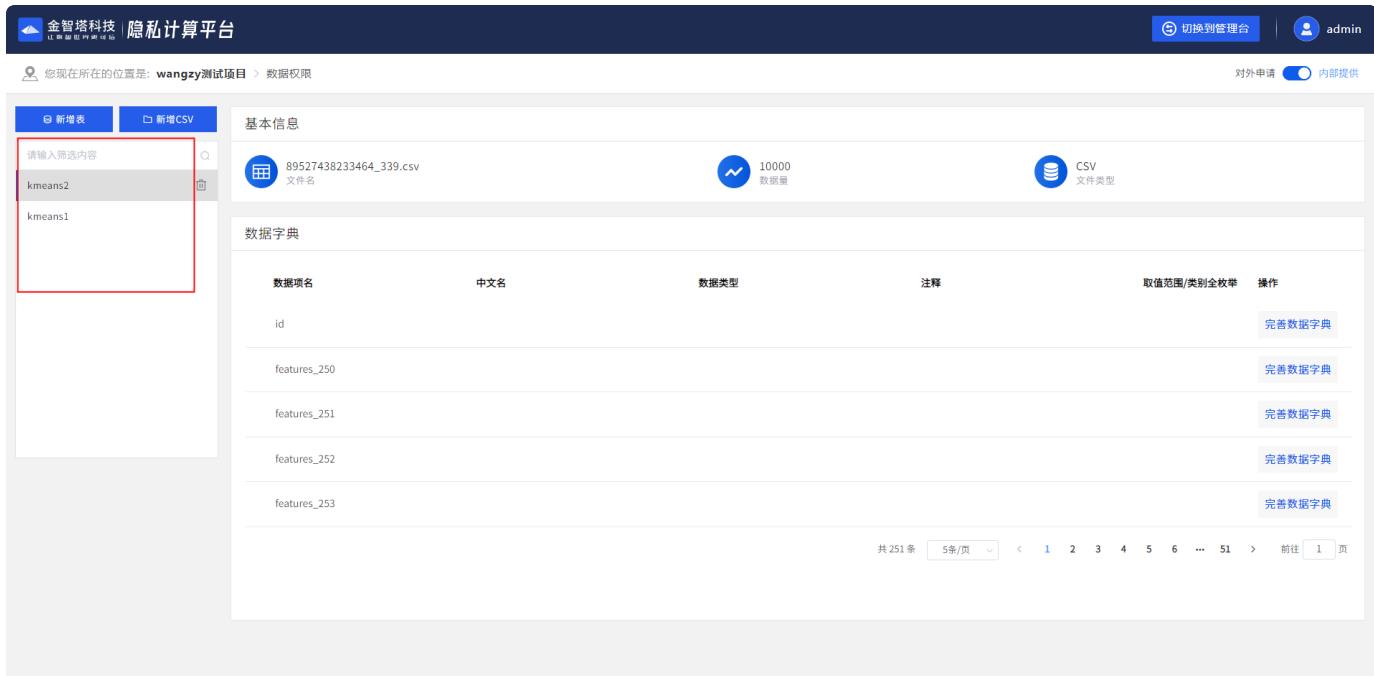


图60 数据目录管理

4.2.2. 数据申请

用户可以在工作台->数据权限页面中，切换到对外申请模式，可以看到合作方的数据源列表，如下图所示。选择所需的数据源，可进行数据使用的权限申请。

图61 合作方的数据源列表

在向其他节点申请数据使用权限勾选特征时，可全选和反选进行申请。点击提交申请即可向对应节点发出申请需求。

图62 数据申请

在提交申请后，在页面下方可以看到待批复、已通过、已拒绝的特征列表，如下图所示。提交申请后，数据提供方可再数据外供审批页面看到相对应的数据申请记录。

图63 数据申请结果

4.3. 项目编辑-画布使用

4.3.1. 画布

用户可将左侧算法库中的组件拖入画布中，通过拖拉拽的形式将各个组件进行连接，创建一个或多个工作流拓扑图。左上角的按钮可以支持放大、缩小和自适应。右上角可以点击运行，进行任务执行。在任务执行完成后，可在右上角切换模式，进行模型的评估。

图64 画布编辑

4.3.2. 组件

点击画布中的组件，可以在右边栏中进行参数的选择和调整。

1. 右边栏中进行组件参数的选择

2. 右键: 运行到此处、删除

图65 组件编辑

组件在运行时，组件会呈现转圈的形式，点击组件，会在画布下方展示组件的实时日志。

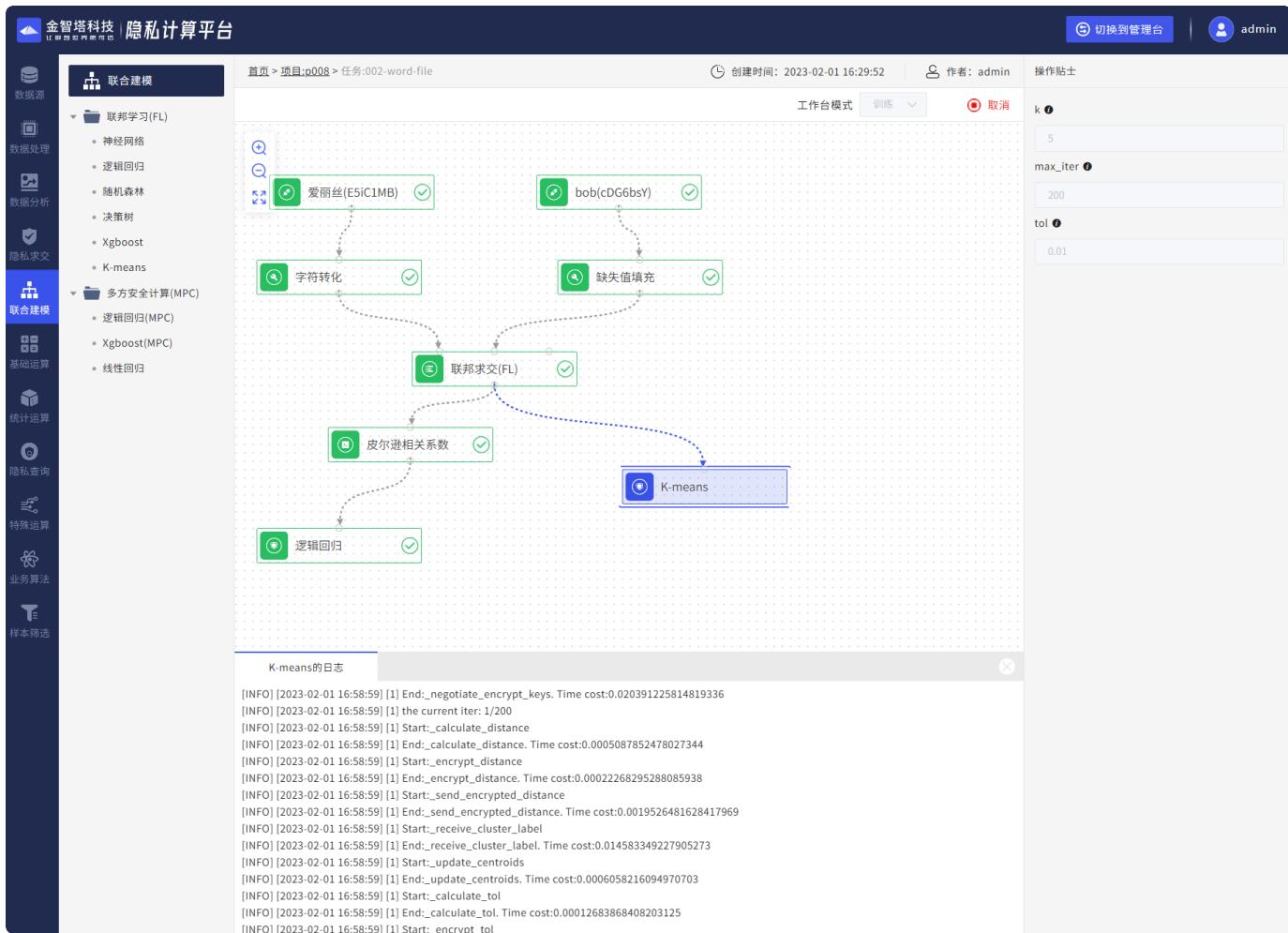


图66 组件实时日志

组件在运行完成后，可右键查看结果详情。

金智塔科技 隐私计算平台

切换到管理台 | admin

隐私求交

联邦求交(FL)
多方安全求交(DH)

创建时间: 2023-02-01 14:25:41 | 作者: admin | 工作台模式 | 训练 | 运行

RSA求交特点:
 1. 非对称加密, 即: PK与SK不是同一个
 2. PK用于加密, SK用于解密
 3. PK决定SK, 但是PK很难算出SK (数学原理: 两个大质数相乘, 积很难因式分解)
 4. 速度慢, 只对少量数据加密

Advanced Encryption Standard (高级加密标准) 特点:
 1. 对称加密
 2. 一个SK扩展成多个子SK, 轮加密

Data Encryption Standard (数据加密标准), 对应算法是DES, 特点:
 1. 对称加密
 2. 同一个SK

加密长度: 256
加密类型: rsa

1. 点击组件, 画布下方展示该组件的实时日志
2. 组件运行完成后, 可右键查看结果详情

联邦求交(FL)的日志

```
[INFO] [2023-02-01 14:50:25] [1] End: __check_data_numeric. Time cost: 0.650520324 / 0.031e-05
[INFO] [2023-02-01 14:50:25] [1] Start: __generate_rsa_key
[INFO] [2023-02-01 14:50:25] [1] End: __generate_rsa_key. Time cost: 0.24627470970153809
[INFO] [2023-02-01 14:50:25] [1] Start: __send_rsa_keys
[INFO] [2023-02-01 14:50:25] [1] End: __send_rsa_keys. Time cost: 0.012724161148071289
[INFO] [2023-02-01 14:50:25] [1] Start: __receive_rsa_keys
[INFO] [2023-02-01 14:50:26] [1] End: __receive_rsa_keys. Time cost: 0.36165881156921387
[INFO] [2023-02-01 14:50:26] [1] Start: __generate_raw_key_and_iv
[INFO] [2023-02-01 14:50:26] [1] End: __generate_raw_key_and_iv. Time cost: 0.0003638267517089844
[INFO] [2023-02-01 14:50:26] [1] Start: __send_keys
[INFO] [2023-02-01 14:50:26] [1] End: __send_keys_to
[INFO] [2023-02-01 14:50:26] [1] End: __send_keys_to. Time cost: 0.004330635070800781
[INFO] [2023-02-01 14:50:26] [1] Start: __send_keys
[INFO] [2023-02-01 14:50:26] [1] End: __send_keys. Time cost: 0.004828691482543945
[INFO] [2023-02-01 14:50:26] [1] Start: __receive_keys
[INFO] [2023-02-01 14:50:26] [1] End: __receive_keys. Time cost: 0.0016553401947021484
[INFO] [2023-02-01 14:50:26] [1] Start: __generate_shared_key
```

图67 组件运行结果

点击组件右键中的数据审计, 可以查看数据中的每个字段的统计信息。

1. 节点列表

2. 特征列表

3. 统计信息

ID	account_age	duration_month	credit_history	purpose	credit_amount	savings_account	present_employment
bob(2DBaKLP)	爱丽丝(u3vJsi4)						

各分区样本数	
等频	216
等宽	72
等频	224
等宽	115
等频	57
等宽	86
等频	87

72.00 最大值	4.00 最小值	18.00 中位数	20.90 平均值
145.42 方差	12.06 标准差	0.58 离散系数	0% 缺失率
0.000001 正态检验P值	- IV	- 相关性系数	- 协方差

图68 组件数据审计

4.3.3. 连接桩

组件的输入桩只能连接一条连线，输出桩可以连接多条连线。其中隐私求交组件的输入桩会随着连线的增加自动增加。

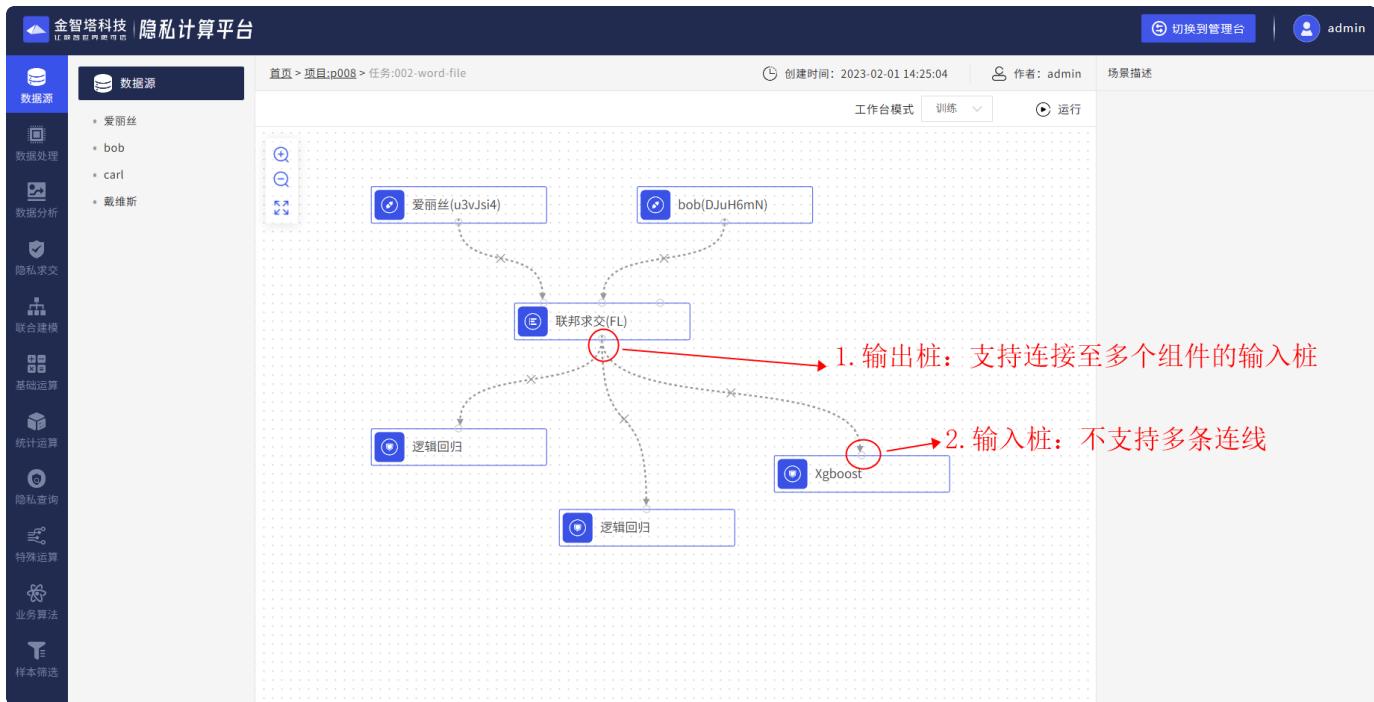


图69 连接桩说明

4.3.4. 连线

支持对连线进行删除。



图70 连线说明

4.3.5. 运行任务

4.3.5.1. 运行

用户可以点击画布右上角的运行按钮，运行画布中所有的工作流拓扑图。

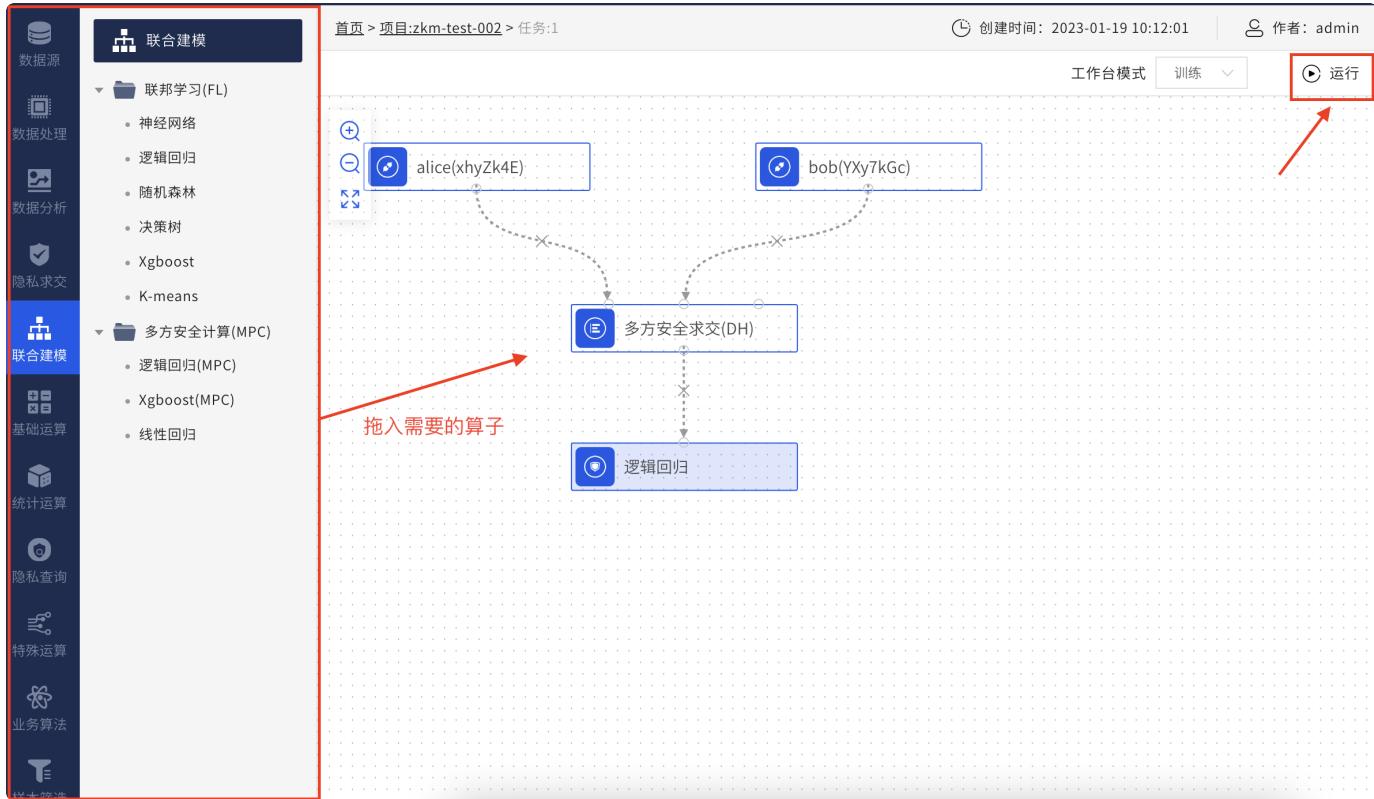


图71 画布的运行

4.3.5.2. 运行到此处

工作流中的每个模块，可通过右击选择运行到此处，进行运行到此模块的操作。此时整个任务只执行该模块及该模块依赖的所有上游模块，该模块的下游模块均不执行。

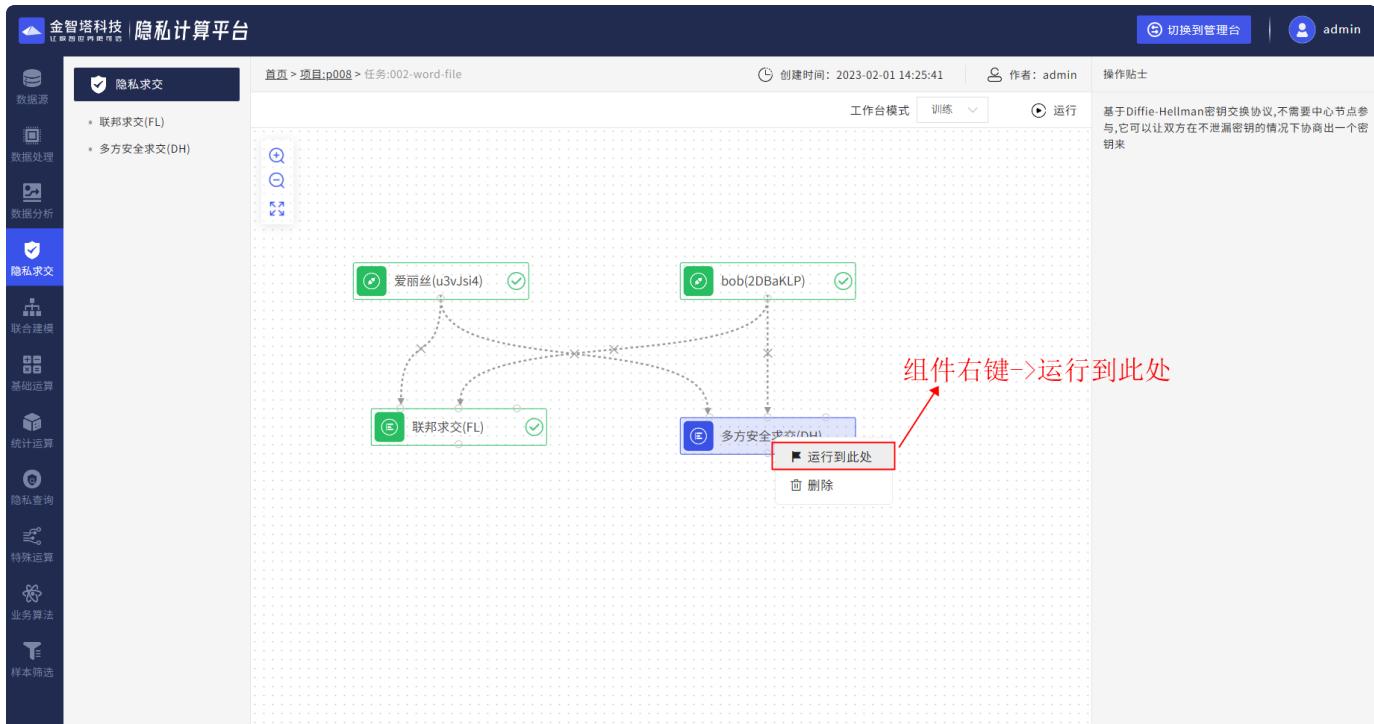


图72 组件的运行到此处

4.3.5.3. 运行此组件

在前面的流程均已执行完成的情况下，点击运行此组件，将只运行当前这一个组件，其后的组件都不执行。此功能只支持部分组件。

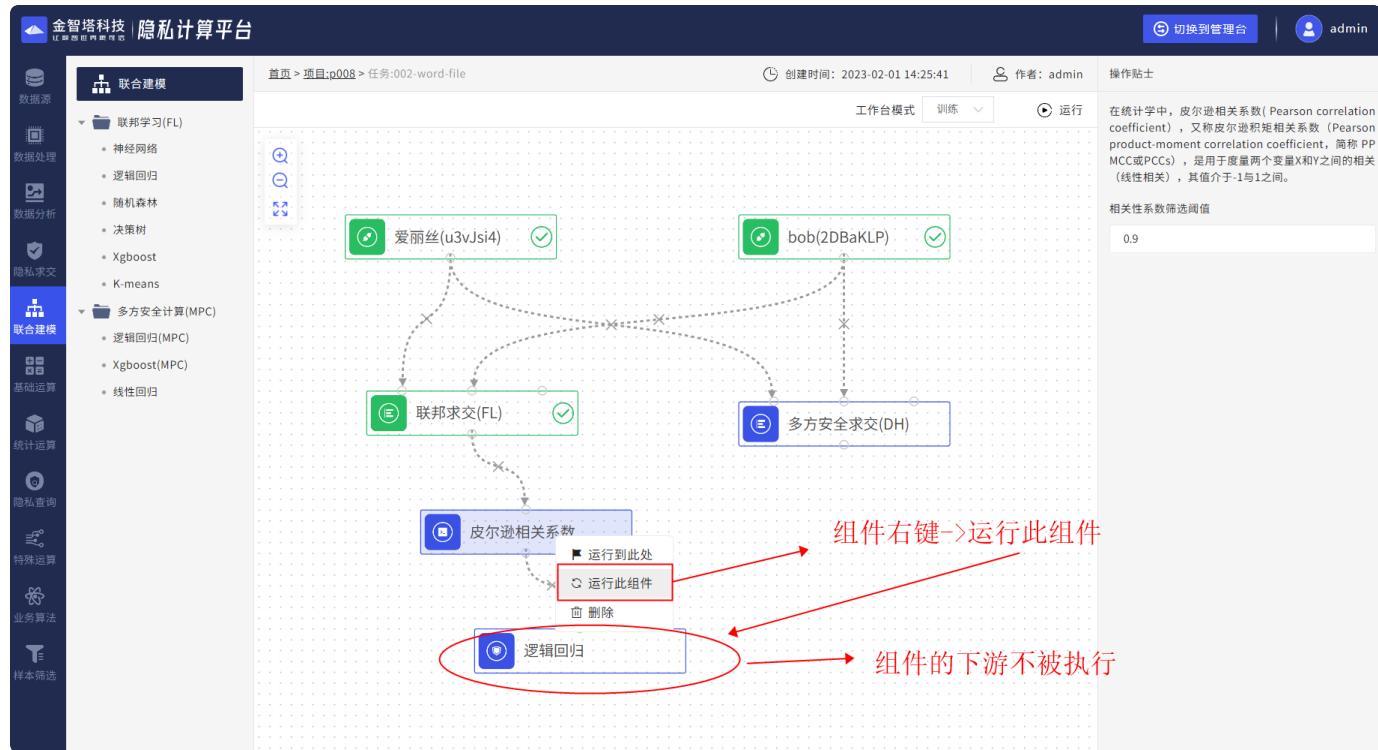


图73 组件的运行此组件

4.4.5 取消任务

4.4.5.1. 取消画布

点击取消，可暂停整个工作流。暂停后，各节点均对该任务中待运行的组件进行暂停操作，并对任务的状态进行保存。

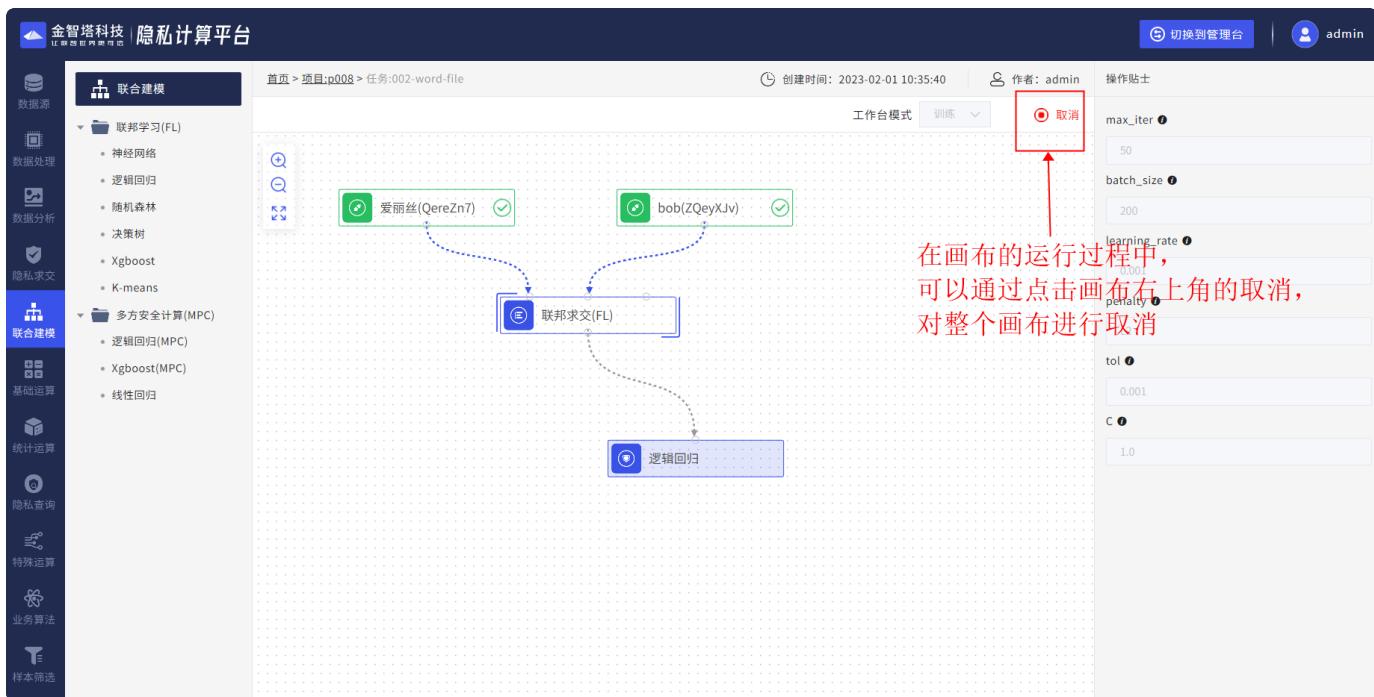


图74 画布的取消

4.4.5.2. 取消组件

工作流中的每个模块，可通过右击选择取消运行，进行此组件任务运行的取消。同时画布上非该组件的上下游组件，其执行逻辑不会受到影响。

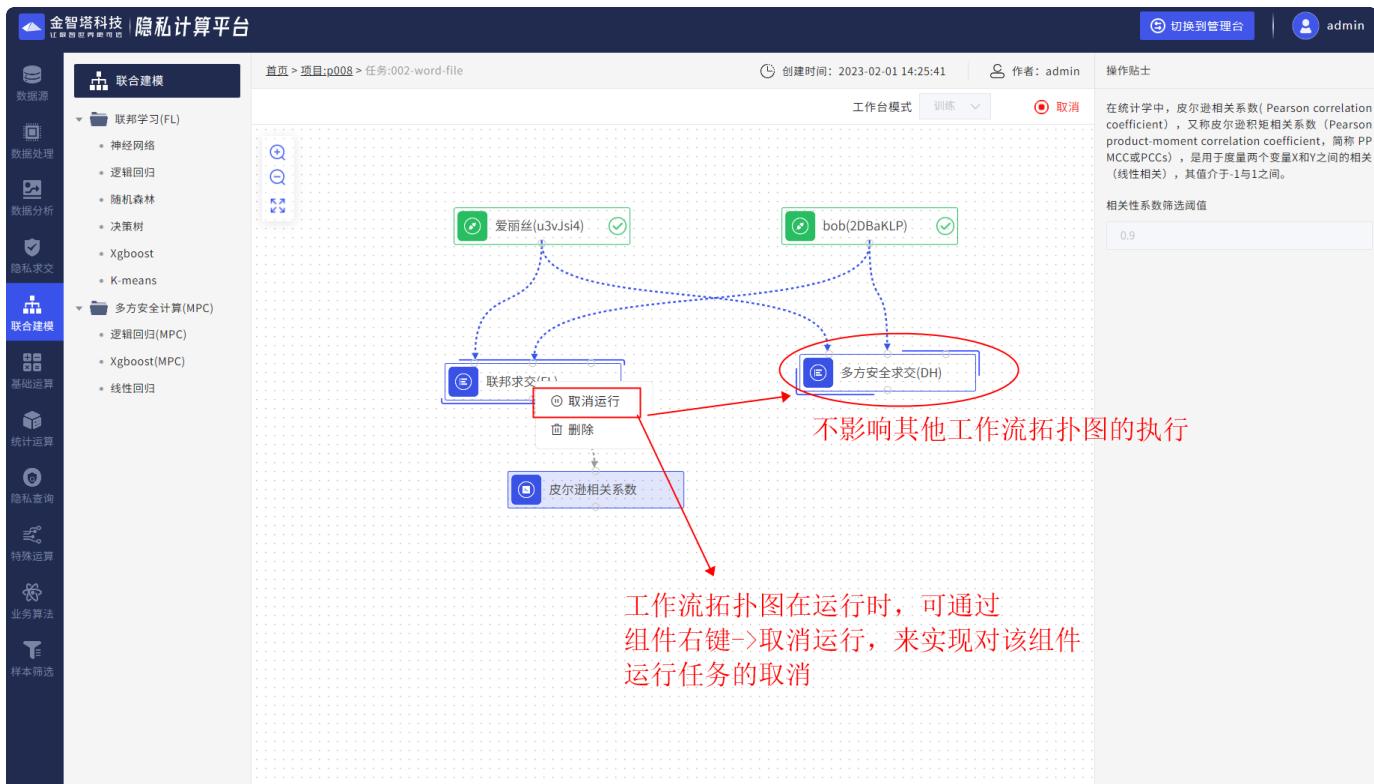


图75 组件的取消运行

4.4. 项目编辑-组件说明

4.4.1. 数据源

用户可选择数据源中的合作方节点，并选择其提供的数据集和字段，进行后续计算。



图76 数据源

4.4.2. 数据处理

在现实的建模过程中，数据常常存在各种问题，数据存在不完全的、有噪声的、不一致的等各种情况。而这些带有错误信息的数据会对模型造成不利的影响。因此在获取到数据源后，需要对数据进行处理，其目的是对各种脏数据进行对应方式的处理，得到标准的、干净的、连续的数据，提供给数据统计、数据挖掘等使用。平台的数据处理组件主要包含预处理（字符转化、稀疏样本删除、缺失特征删除、缺失值填充）、特征工程（特征分箱、卡方分箱、类别编码、标准化）、自定义加工（Sql 编辑、Python 编辑）三大部分。

4.4.2.1. 字符转化

a) 组件说明

支持单方处理、多方处理（组件位于隐私求交组件的下游），可以将特征的数据格式转化为int、float、str、datetime、bool五种类型。

b) 画布编辑

在数据处理模块拖入字符转化组件，可选择批量操作或对单个特征进行目标类型的转化。

The screenshot shows the 'Data Processing' library with the 'Character Conversion' component selected. A tooltip indicates: '1、从数据处理库中拖入字符转化组件' (Drag the character conversion component from the data processing library). On the right, a modal window titled '批量操作' (Batch Operation) shows a dropdown menu for selecting data types: 'int', 'float', 'str', 'datetime', and 'bool'. A red box highlights the '批量操作' button. Another red box highlights the dropdown menu. A third red box highlights the '目标类型' (Target Type) column in the table on the right, which lists various feature names and their current types.

图77 字符转化

The screenshot shows the 'Data Processing' library with the 'Character Conversion' component selected. A tooltip indicates: '1. 从组件库中选择组件' (Select the component from the component library). A red box highlights the 'Character Conversion' component icon. Another red box highlights the '参数栏' (Parameter Bar) on the right, which contains a dropdown menu for selecting data types: 'int', 'float', 'str', 'datetime', and 'bool'. A third red box highlights the '目标类型' (Target Type) column in the table on the right, which lists various feature names and their current types.

图78 字符转化的多方处理

c) 运行结果

- 在组件执行完成后，可通过右键查看更多详情。

操作贴士	操作	值	类型	状态
alice(Xd...	grade	未定义	未定义	
alice(Xd...	subGrade	未定义	未定义	
alice(Xd...	employ...	未定义	未定义	
alice(Xd...	employ...	未定义	未定义	
alice(Xd...	homeO...	未定义	未定义	
alice(Xd...	annual...	未定义	未定义	
alice(Xd...	verificat...	未定义	未定义	
alice(Xd...	issueDate	未定义	datetime	
alice(Xd...	purpose	未定义	未定义	
alice(Xd...	postCode	未定义	未定义	
alice(Xd...	regionC...	未定义	未定义	
alice(Xd...	dti	未定义	未定义	
alice(Xd...	delinqu...	未定义	未定义	
alice(Xd...	ficoRan...	未定义	未定义	
alice(Xd...	ficoRan...	未定义	未定义	
alice(Xd...	openAcc	未定义	未定义	
alice(Xd...	pubRec	未定义	未定义	
alice(Xd...	pubRec...	未定义	未定义	

图79 字符转化组件运行结果右键

- 点击查看数据可查看哪些特征进行了字符类型转化。

图80 字符转化组件运行结果右键查看数据

4.4.2.2. 稀疏样本删除

a) 组件说明

在建模过程中，如果样本有大部分特征缺失时，该样本属于异常样本，会影响到模型的训练，需要将其删除。

该组件支持单方处理、多方处理（组件位于隐私求交组件的下游），可设置参数缺失率，当样本的缺失率大于设定的参数（阈值）的时候，将其删除。

b) 画布编辑

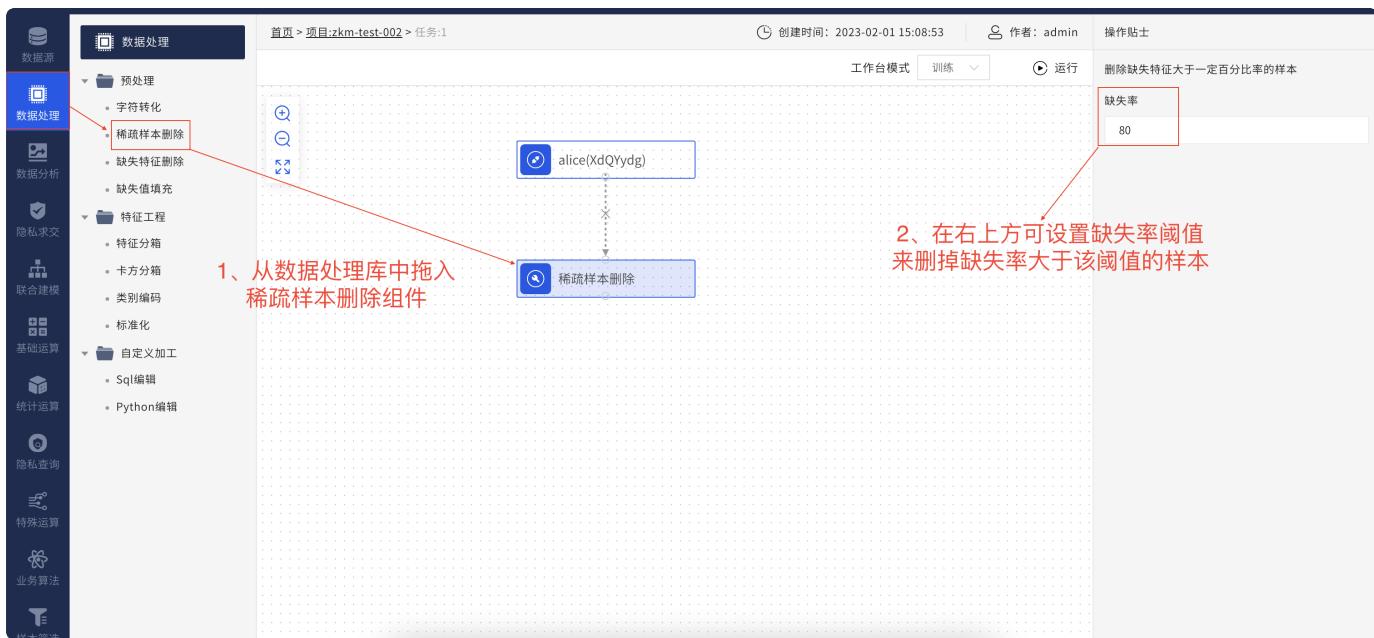


图81 稀疏样本删除

c) 运行结果

- 在组件执行完成后，可通过右键查看更多详情：同4.4.2.1中数据处理-字符转化c）。
- 点击查看数据可查看删除稀疏样本后数据的维度大小。

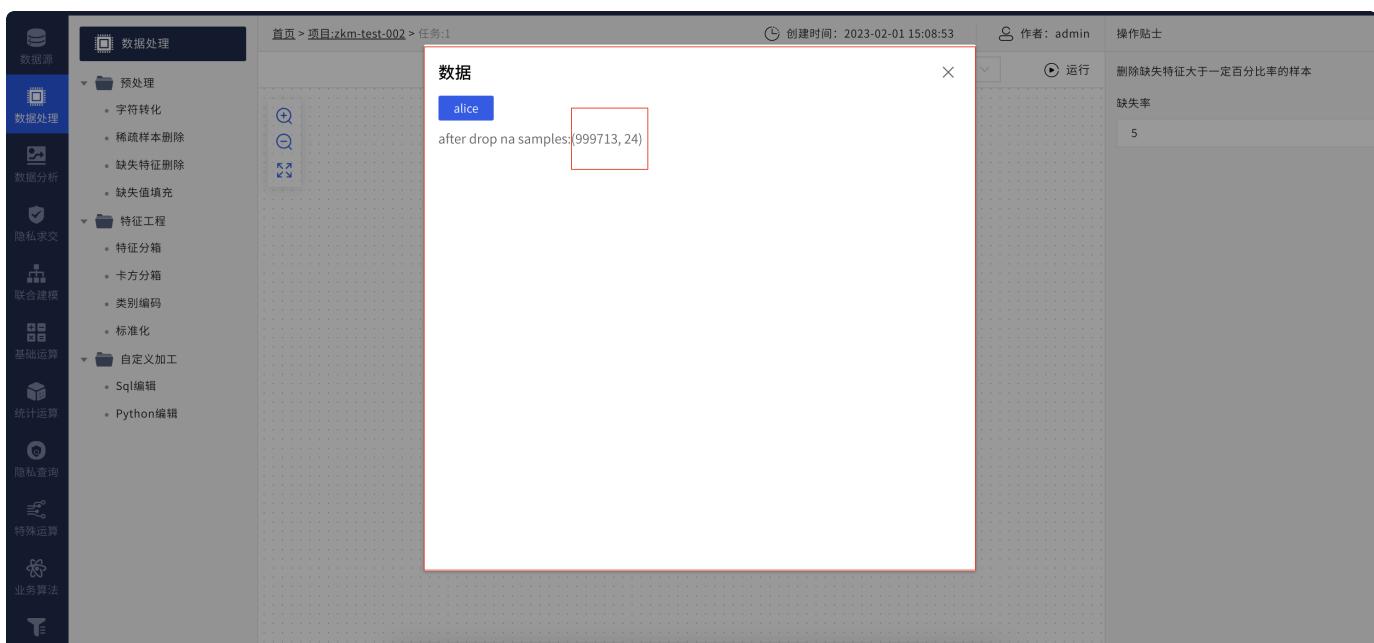


图82 稀疏样本删除组件运行结果右键查看数据

4.4.2.3. 缺失特征删除

a) 组件说明

在建模过程中，如果某个特征有大部分缺失，会影响到模型的训练，造成结果的偏差，可认为该特征属于无效特征，应该将其删除。

该组件支持单方处理、多方处理（组件位于隐私求交组件的下游），当特征的缺失率大于设定的删除临界比例时，可认为该特征是无效特征，将其删除。

b) 画布编辑

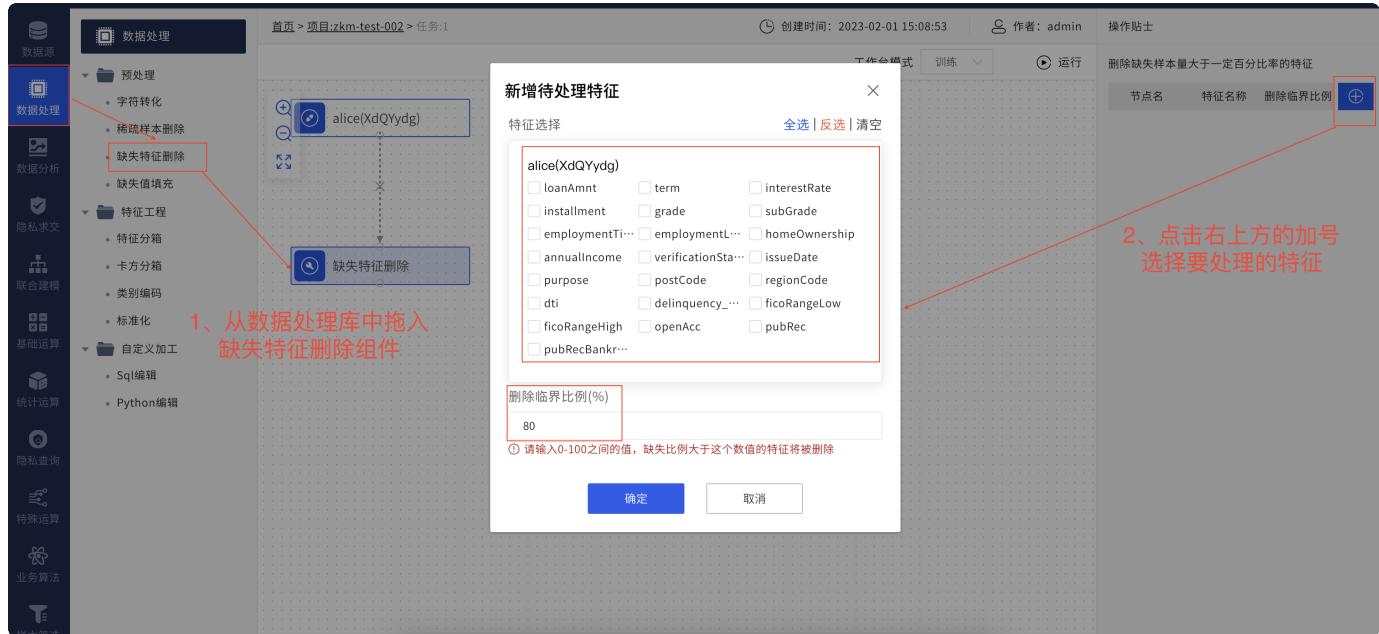


图83 缺失特征删除

c) 运行结果

- 在组件执行完成后，可通过右键查看更多详情：同4.4.2.1数据处理-字符转化c）。
- 点击查看数据可查看哪些特征被删除。

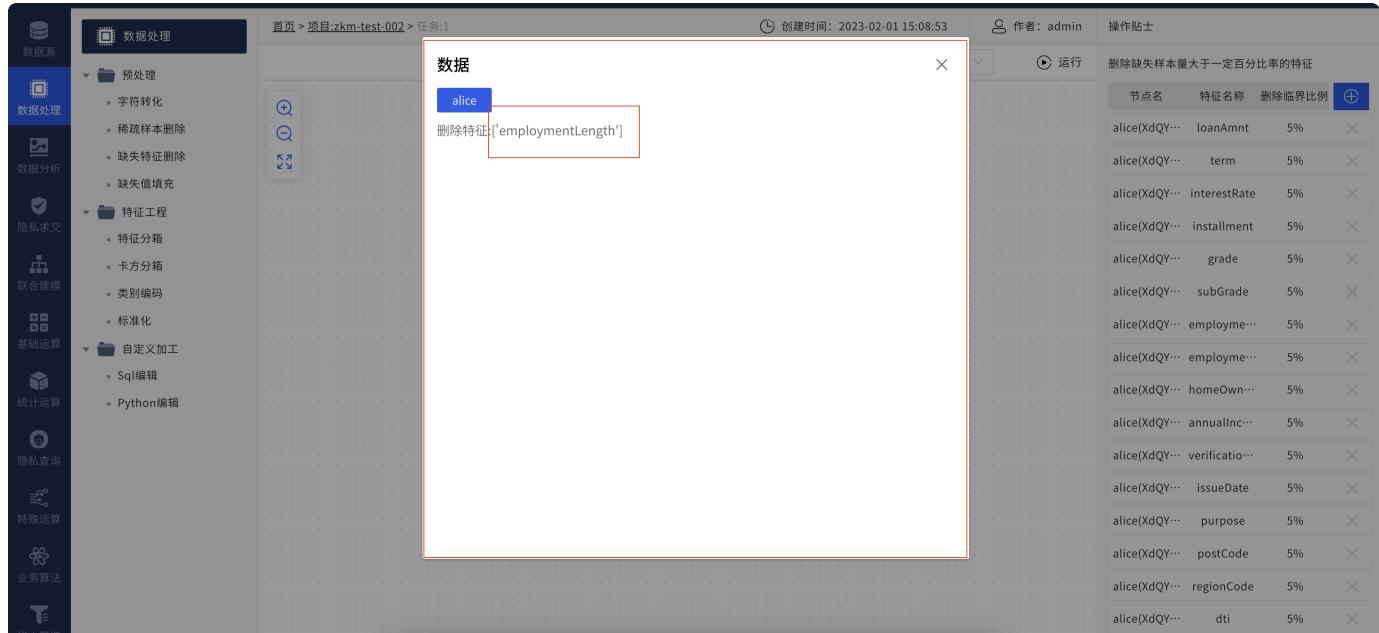


图84 缺失特征删除组件运行结果右键查看数据

4.4.2.3. 缺失值填充

a) 组件说明

当特征缺失率较小时，可以对特征进行缺失值填充，填充方式有：**均值**、**中位数**、**其他**三种方式。该组件支持单方处理、多方处理（组件位于隐私求交组件的下游）。

b) 画布编辑

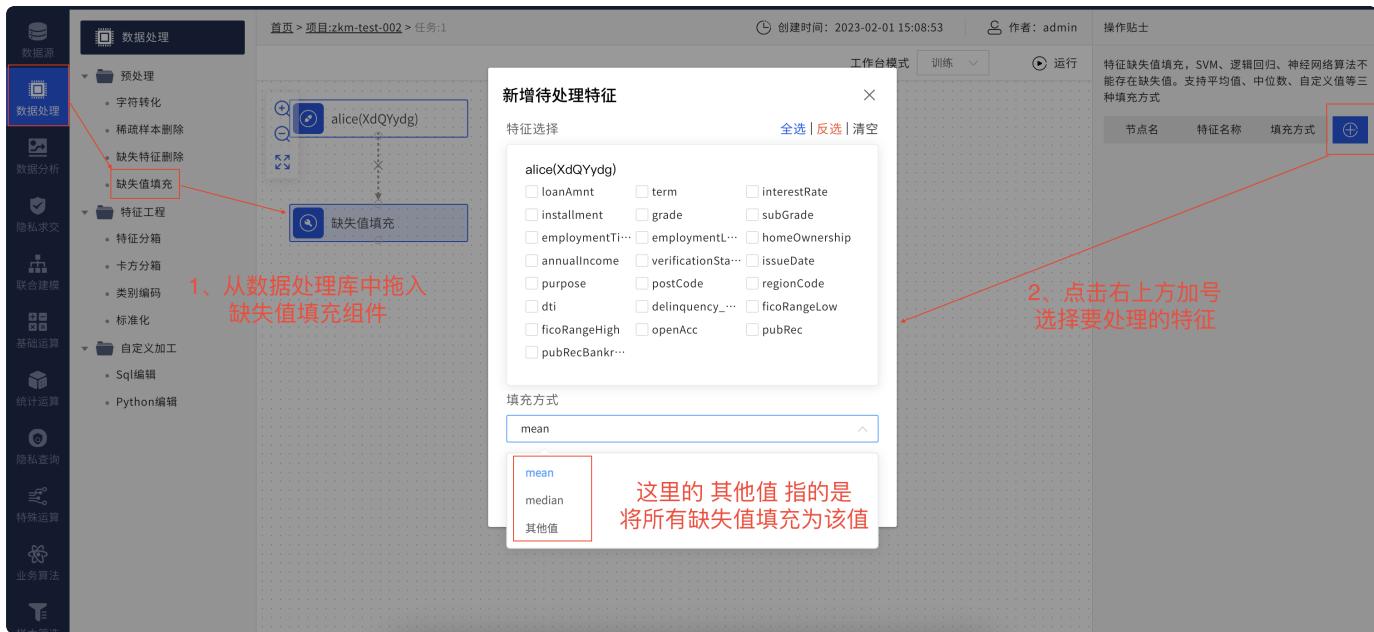


图85 缺失值填充

c) 运行结果

- 在组件执行完成后，可通过右键查看更多详情：同4.4.2.1数据处理-字符转化c）。
- 点击查看数据可查看哪些特征进行了缺失值填充，以及选中的特征中哪些是没有缺失值的。

节点名	特征名称	填充方式
金智塔科技...	loanAmnt	median
金智塔科技...	term	median
金智塔科技...	interestRate	median
金智塔科技...	installment	median
金智塔科技...	grade	median
金智塔科技...	subGrade	median
金智塔科技...	employmentLength	median
金智塔科技...	annualIncome	median
金智塔科技...	verificationStatus	median
金智塔科技...	issueDate	median
金智塔科技...	purpose	median
金智塔科技...	postCode	median
金智塔科技...	regionCode	median

图86 缺失值填充组件运行结果右键查看数据

4.4.2.4. 特征分箱

a) 组件说明

特征分箱是一种数据预处理技术，用于减少次要观察误差的影响，是一种将多个连续值分组为较少数量的“分箱”的方法。一般在建立分类模型时，需要对连续变量离散化，特征离散化后，模型会更稳定，降低了模型过拟合的风险。比如在建立申请评分卡模型时用logistic作为基模型就需要对连续变量进行离散化，离散化通常采用分箱法。分箱方法常见的有等频分箱（每个区间内包括的值一样多）和等距分箱（每两区间之间的距离是一样的）。

支持单方处理、多方处理（组件位于隐私求交组件的下游）。

b) 画布编辑

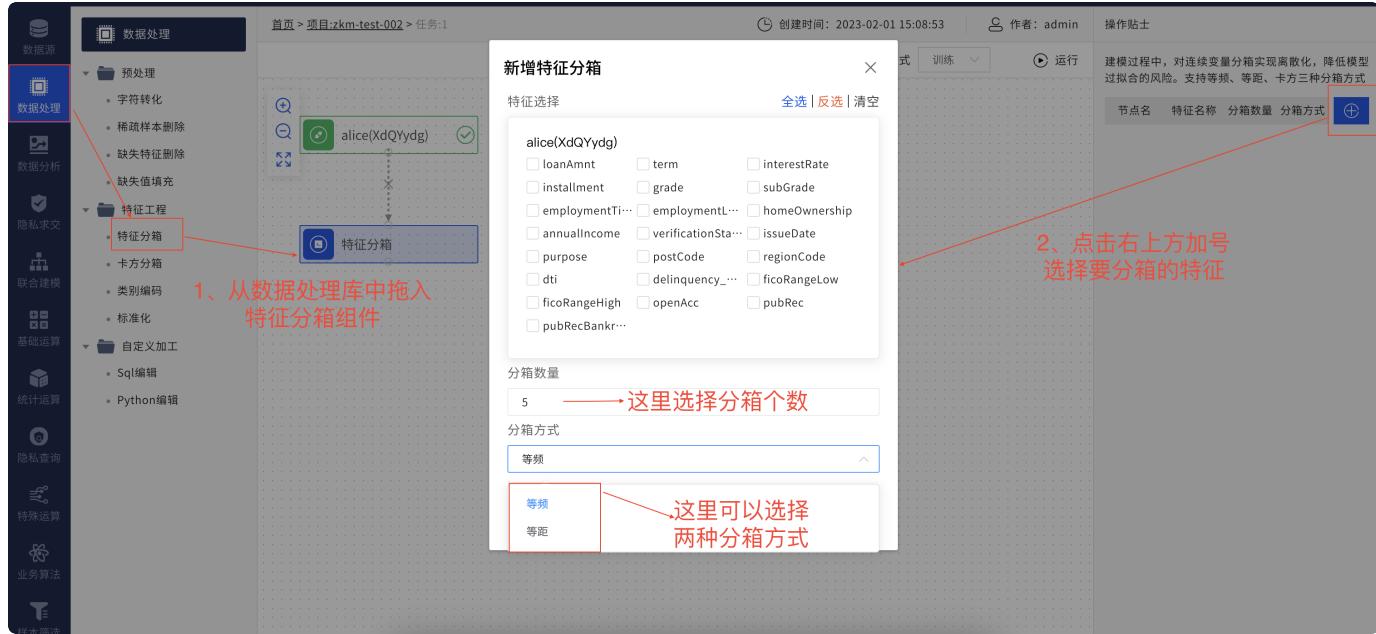


图87 特征分箱

c) 运行结果

- 在组件执行完成后，可通过右键查看更多详情：同4.4.2.1数据处理-字符转化c）。
- 点击查看数据可查看哪些特征进行了特征分箱。

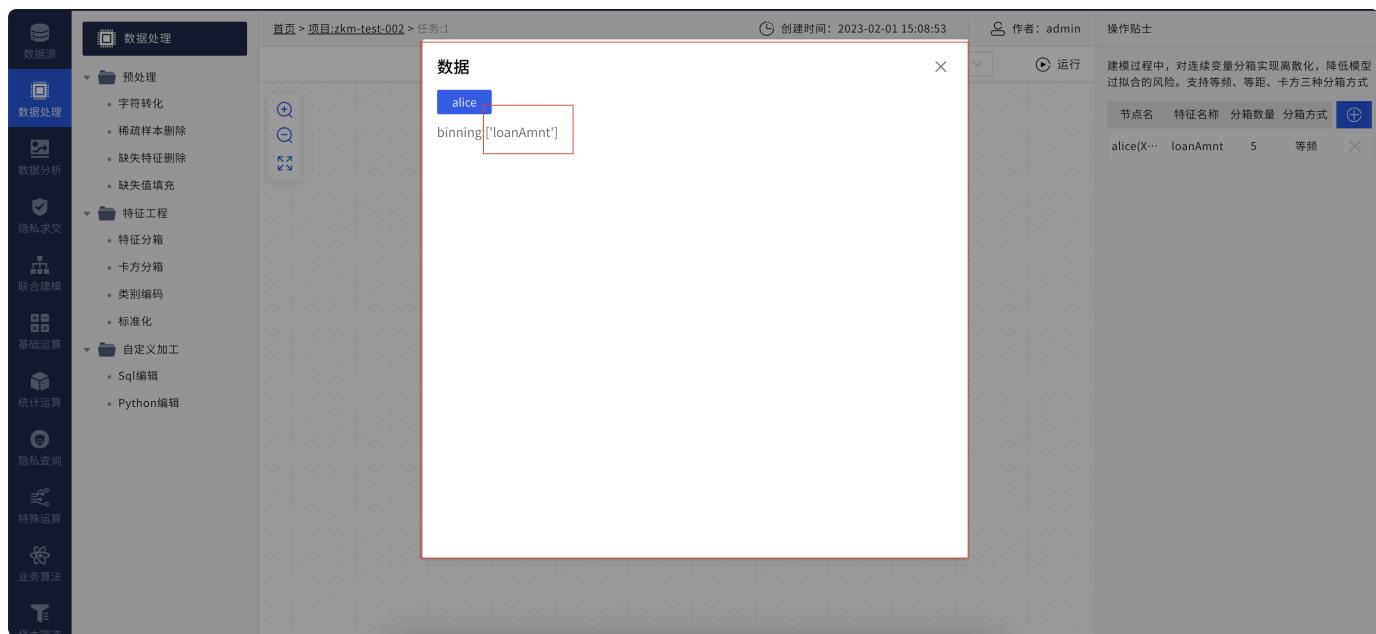


图88 特征分箱组件运行结果右键查看数据

4.4.2.5. 卡方分箱

a) 组件说明

卡方分箱是自底向上的(即基于合并的)数据离散化方法。它依赖于卡方检验：具有最小卡方值的相邻区间合并在一起,直到满足确定的停止准则。基本思想：对于精确的离散化，相对类频率在一个区间内

应当完全一致。因此,如果两个相邻的区间具有非常类似的类分布,则这两个区间可以合并;否则,它们应当保持分开。而低卡方值表明它们具有相似的类分布。其实现步骤主要包括以下两个阶段:

- **初始化阶段:**

首先按照属性值的大小进行排序(对于非连续特征,需要先做数值转换,比如转为坏人率,然后排序),然后每个属性值单独作为一组。

- **合并阶段:**

- (1) 对每一对相邻的组,计算卡方值。
- (2) 根据计算的卡方值,对其中最小的一对邻组合并为一组。
- (3) 不断重复(1)、(2),直到分组数达到一定的条件(如最大分组数5)。

b) 画布编辑

卡方分箱组件必须在求交后执行,可设置分箱数量。

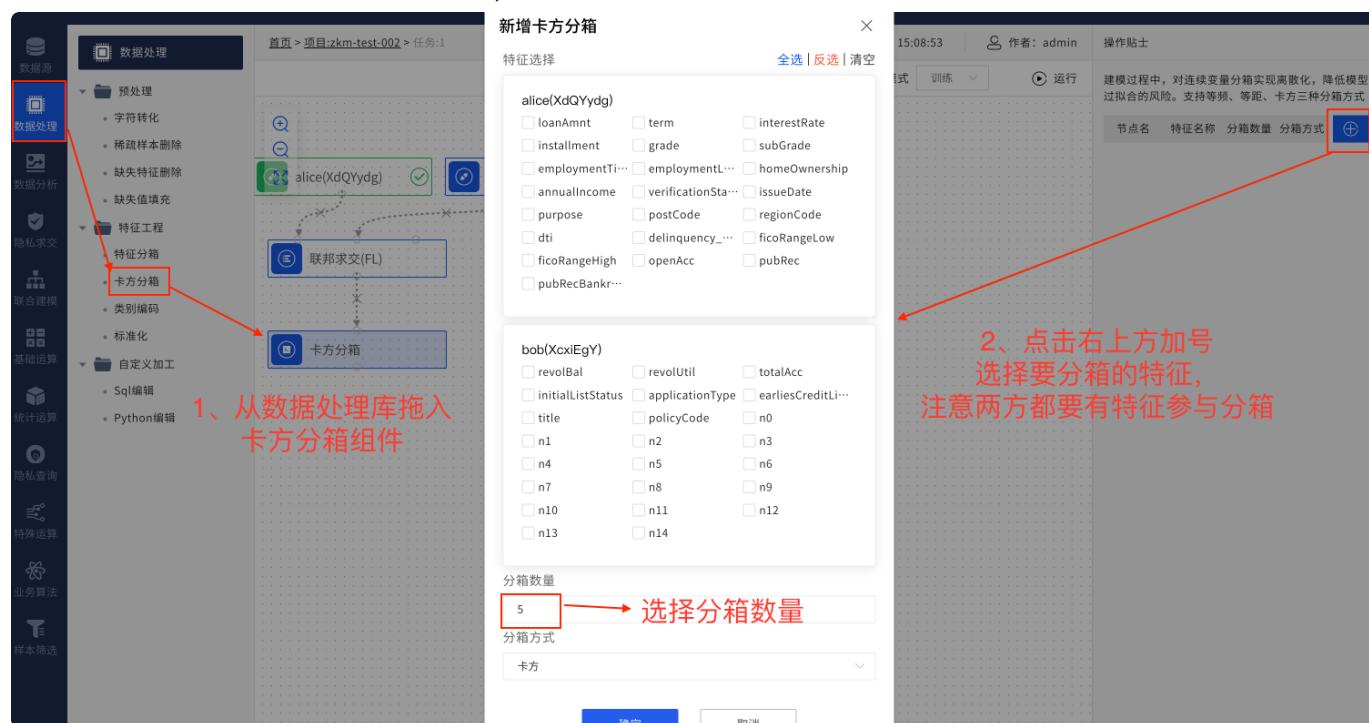


图89 卡方分箱

c) 运行结果

- 在组件执行完成后,可通过右键查看更多详情: 同4.4.2.1数据处理-字符转化c)。
- 点击**查看数据**可查看哪些特征进行了卡方分箱。

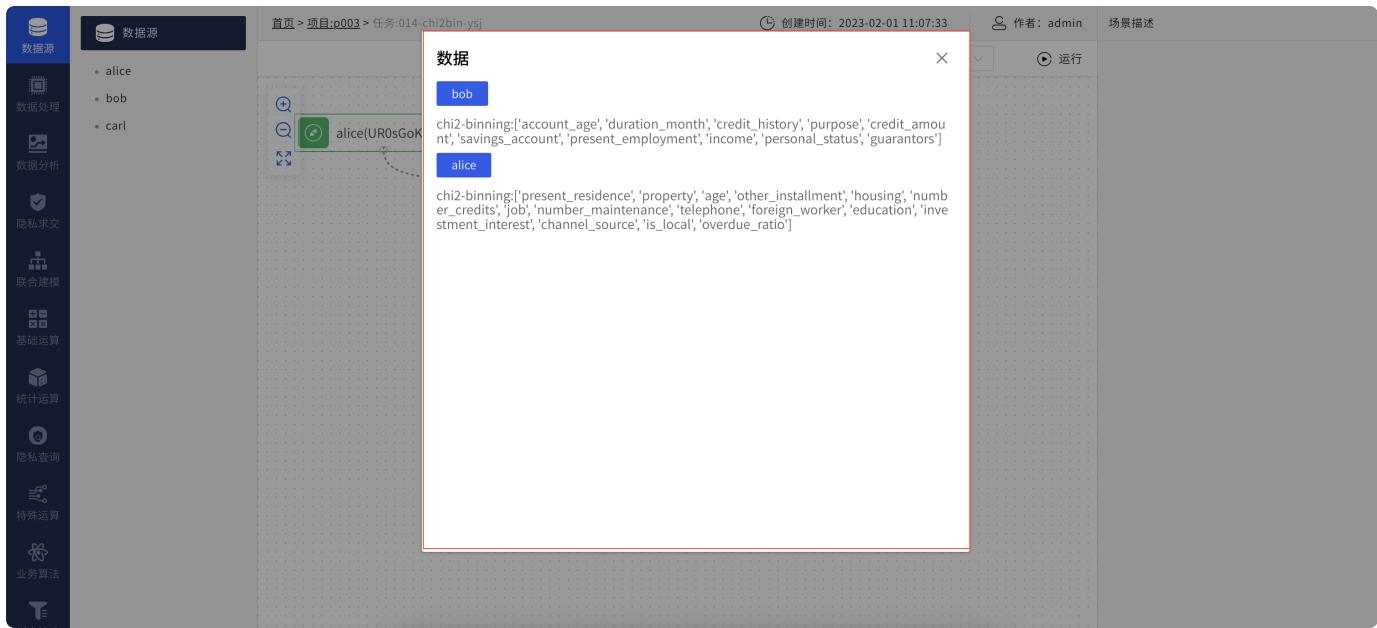


图90 卡方分箱组件运行结果右键查看数据

4.4.2.5. 类别编码

a) 组件说明

大多数机器学习模型只能处理数字。数值（连续、定量）变量是可以在有限或无限区间内取任何值的变量，它们可以很自然地用数字表示，所以可以在模型中直接使用。原始类别变量通常以字符串的形式存在，在传入模型之前需要变换。这里的类别编码指的就是直接统计类别变量的类别数，然后按照阿拉伯数字顺序编码。

该组件支持单方处理、多方处理（组件位于隐私求交组件的下游）。

b) 画布编辑

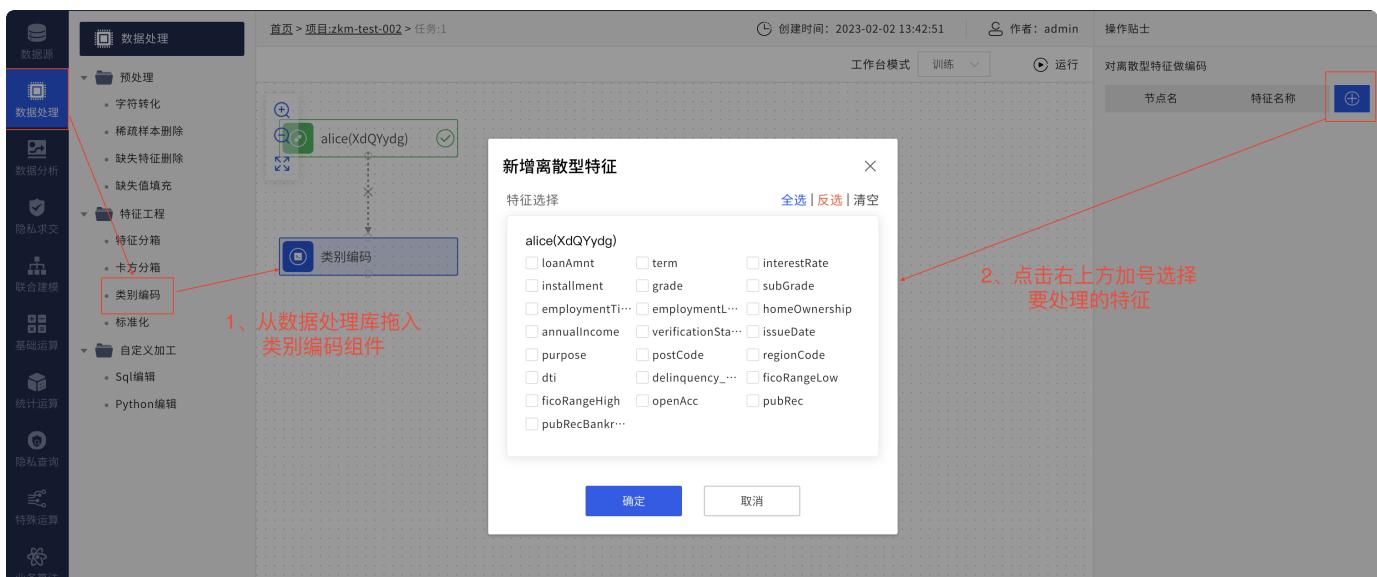


图91 类别编码

c) 运行结果

- 在组件执行完成后，可通过右键查看更多详情：同4.4.2.1数据处理-字符转化c）。

- 点击查看数据可查看哪些特征进行了特征编码。

图92 类别编码组件运行结果右键查看数据

4.4.2.6. 标准化

a) 组件说明

数据的标准化 (normalization) 是将数据按比例缩放，使之落入一个较小的特定区间。在某些比较和评价的指标处理中经常会用到，去除数据的单位限制，将其转化为无量纲的纯数值，便于不同单位或量级的指标能够进行比较和加权。

在机器学习算法的目标函数(例如SVM的RBF内核或线性模型的l1和l2正则化)，许多学习算法中目标函数的基础都是假设所有的特征都是零均值并且具有同一阶数上的方差。如果某个特征的方差比其他特征大几个数量级，那么它就会在学习算法中占据主导位置，导致学习器并不能像我们期望的那样，从其他特征中学习。举一个简单的例子，在KNN中，我们需要计算待分类点与所有实例点的距离。假设每个实例点 (instance) 由n个features构成。如果我们选用的距离度量为欧式距离，如果数据预先没有经过标准化，那么那些绝对值大的features在欧式距离计算的时候起了决定性作用。标准化分为**均值方差标准化 (z_score标准化)** 和**最大值最小值归一化 (min_max归一化)**。

z_score 标准化计算公式为：

对序列 x_1, x_2, \dots, x_n 进行变换：

$$y_i = \frac{x_i - \bar{x}}{s}, \text{ 这里 } \bar{x} = \frac{1}{n} \sum_{i=1}^n x_i, \quad s = \sqrt{\frac{1}{n-1} \sum_{i=1}^n (x_i - \bar{x})^2}$$

则新序列 y_1, y_2, \dots, y_n 的均值为 0，而方差为 1，且无量纲。

min_max 归一化计算公式为：

对序列 x_1, x_2, \dots, x_n 进行变换：

$$y_i = \frac{x_i - \min_{1 \leq j \leq n} \{x_j\}}{\max_{1 \leq j \leq n} \{x_j\} - \min_{1 \leq j \leq n} \{x_j\}}$$

则新序列 $y_1, y_2, \dots, y_n \in [0, 1]$ 且无量纲。一般的数据需要时都可以考虑先进行规范化处理。

该组件支持单方处理、多方处理（组件位于隐私求交组件的下游）。

b) 画布编辑



图93 标准化

c) 运行结果

- 在组件执行完成后，可通过右键查看更多详情：同4.4.2.1数据处理-字符转化c）。
- 点击查看数据可查看哪些特征进行了标准化。

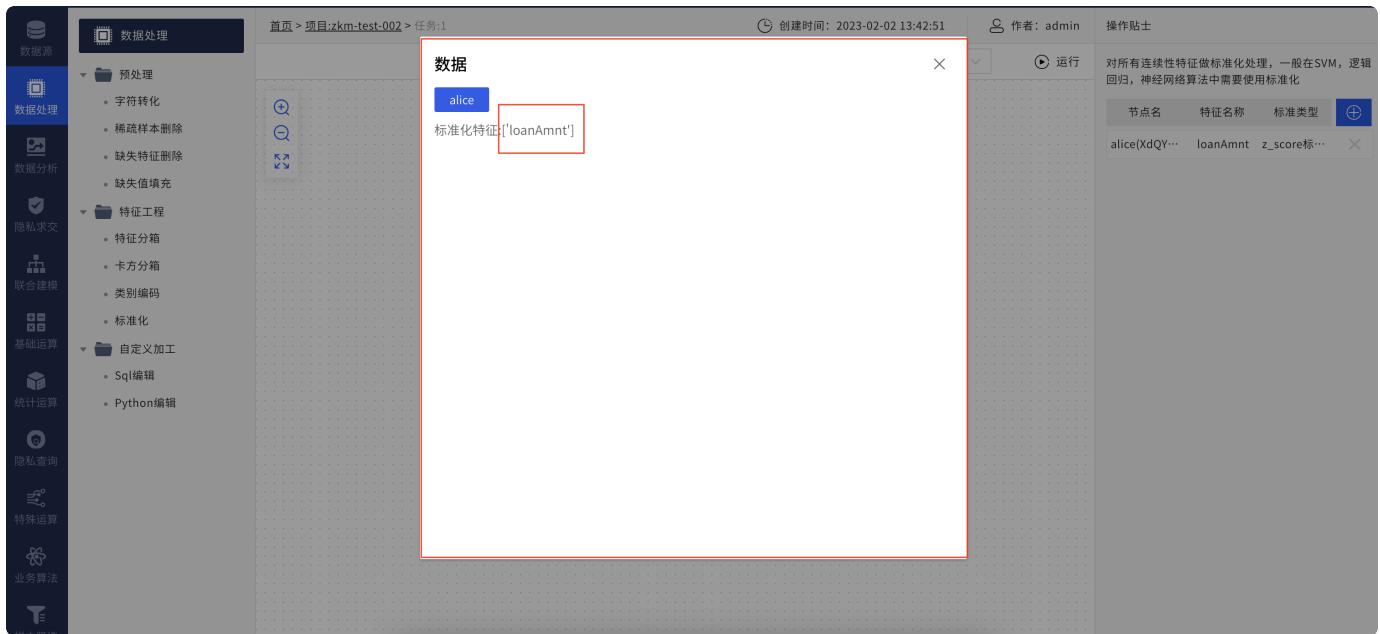


图94 标准化组件运行结果右键查看数据

4.4.2.7. Sql编辑

a) 组件介绍

可通过自定义Sql语句进行数据处理和特征工程（需要上游的组件为数据库形式的数据源节点）。

b) 画布编辑

需要注意的是，Sql编辑的语法需要和上游数据源数据库类型保持一致。

如当数据库为Mysql时，限定返回10条记录数的语法为 `limit 10`；当数据库为DB2时，限定返回10条记录数为`fetch first 10 rows`。



图95 数据源节点提供数据库数据

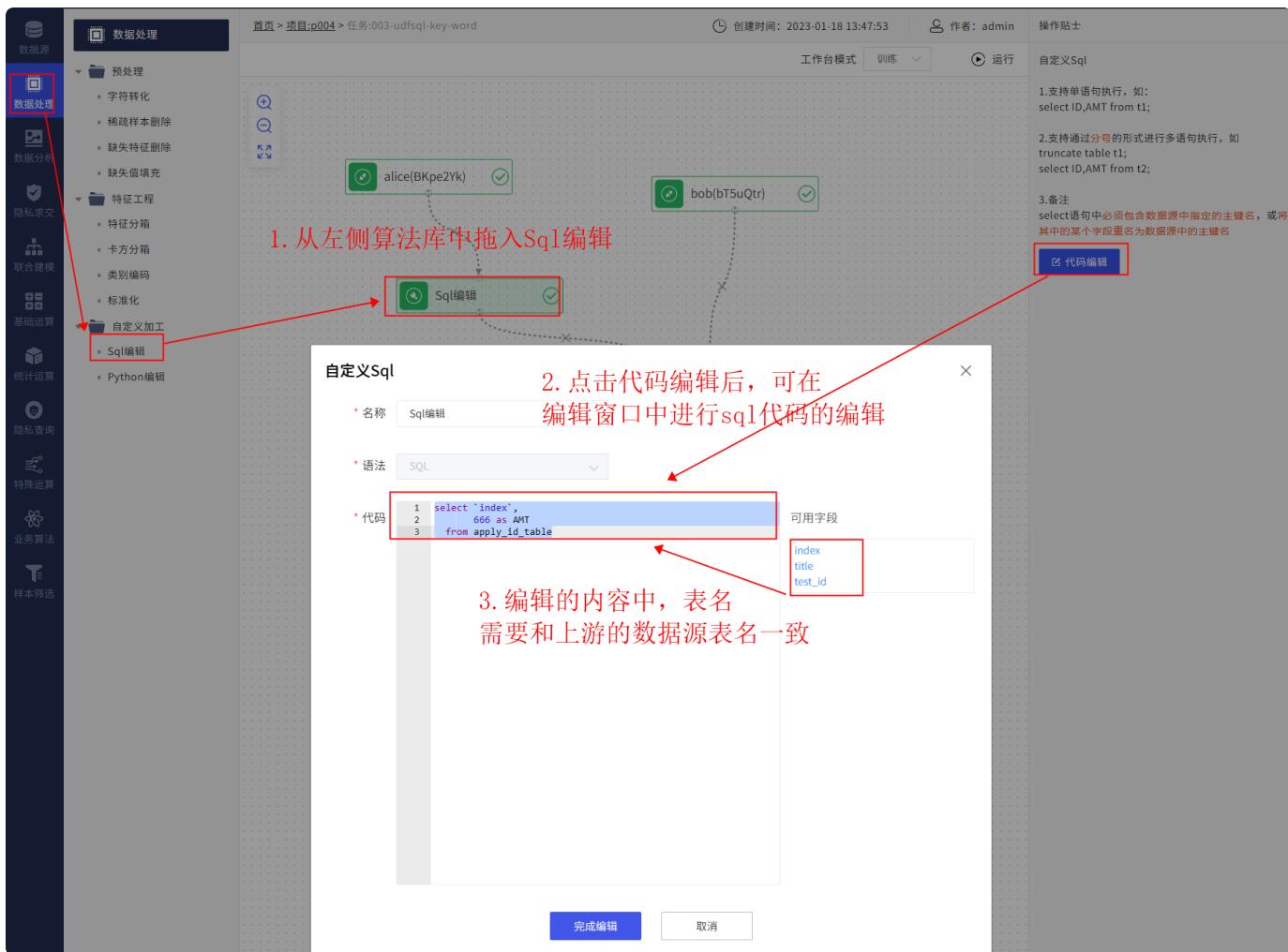


图96 Sql编辑

c) 运行结果

在组件执行完成后，可通过右键点击数据审计，查看处理后的字段的统计数据。

可以在数据审计中查看处理后的字段的统计信息

	最大值	最小值	中位数	平均值
方差	0	0	0	0%
正态检验P值	NaN	-	-	协方差

图97 Sql编辑组件的数据审计

4.4.2.8. Python编辑

a) 组件介绍

可通过自定义Python语句进行数据处理和特征工程，注意：输入输出均是dataframe。

支持单方处理、多方处理（组件位于隐私求交组件的下游）。

b) 画布编辑

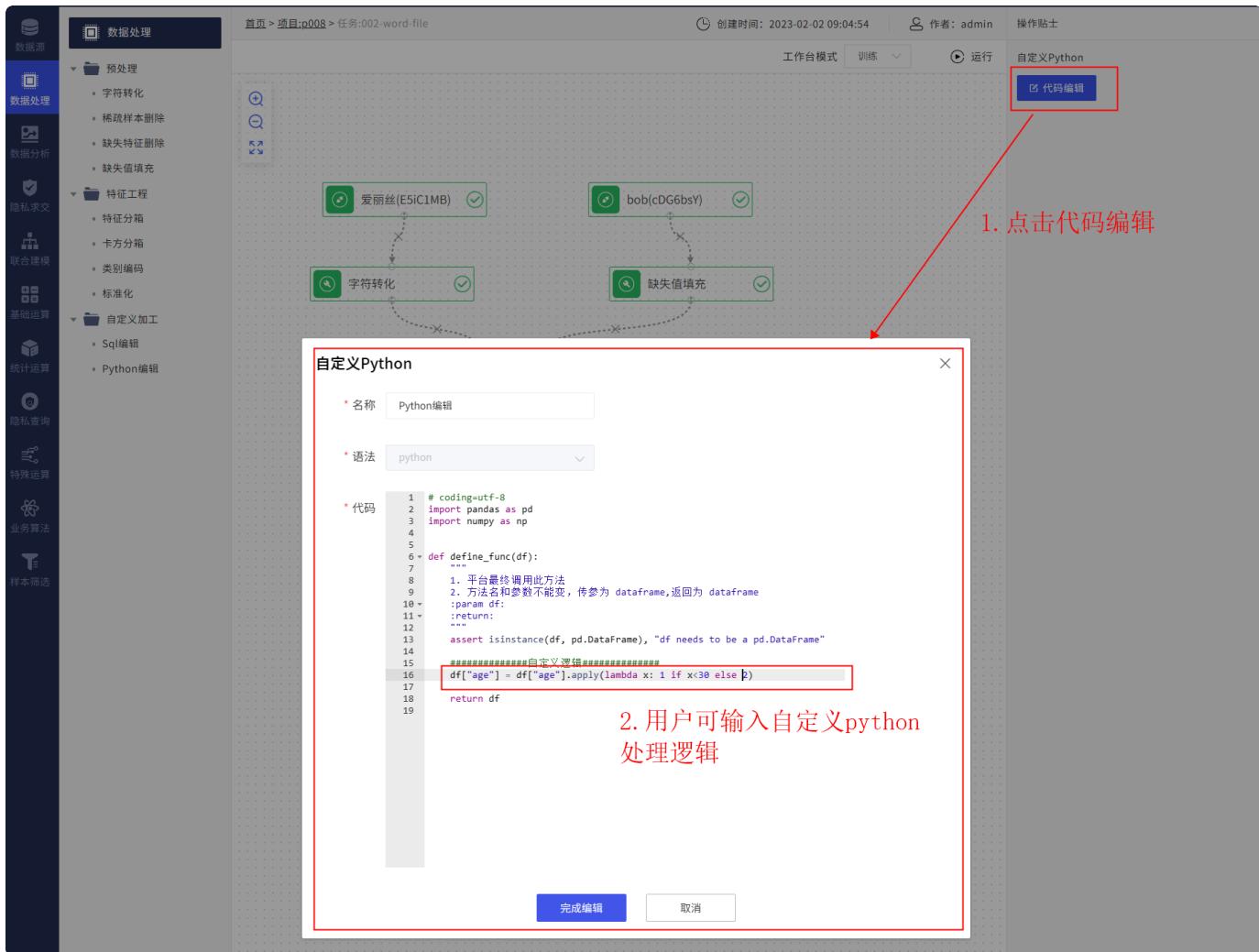


图98 Python编辑

c) 运行结果

在组件执行完成后，可通过右键点击数据审计，查看处理后的字段的统计数据。



图99 Python编辑组件的数据审计

4.4.3. 数据分析

特征筛选在特征工程中，是非常重要的一个环节，其目标是**寻找最优特征子集**。特征筛选能剔除不相关(irrelevant)或冗余(redundant)的特征，从而达到减少特征个数，**提高模型精确度，减少运行时间的目的**。另一方面，选取出真正相关的特征简化模型，能协助理解数据产生的过程。我们常能听到“数据和特征决定了机器学习的上限，而模型和算法只是逼近这个上限而已”，由此可见其重要性。平台的特征筛选方法共有4种：**皮尔逊相关系数、信息值IV、特征重要性分析（树模型）、方差膨胀系数VIF**。

4.4.3.1. 皮尔逊相关系数

a) 组件介绍

Pearson相关系数是用来检测两个**连续型变量**之间**线性相关**的程度。对于两个变量X、Y，其皮尔逊相关系数计算公式如下：

$$\rho_{X,Y} = \frac{\text{cov}(X, Y)}{\sigma_X \sigma_Y} = \frac{E[(X - E(X))(Y - E(Y))]}{\sigma_X \sigma_Y} = \frac{E(XY) - E(X)E(Y)}{\sqrt{E(X^2) - E^2(X)}\sqrt{E(Y^2) - E^2(Y)}}$$

cov为协方差， σ 为标准差。

Pearson相关系数的取值范围为[-1,1]，正值表示正相关，负值表示负相关，绝对值越大表示线性相关程度越高。用户可通过设置**相关性系数筛选阈值**来进行特征筛选，当两个特征的皮尔逊相关系数大于该阈值时，会去掉其中一个特征，只保留剩下的一个。

b) 画布编辑

该组件需在求交后执行，且**必须先做缺失值处理**。

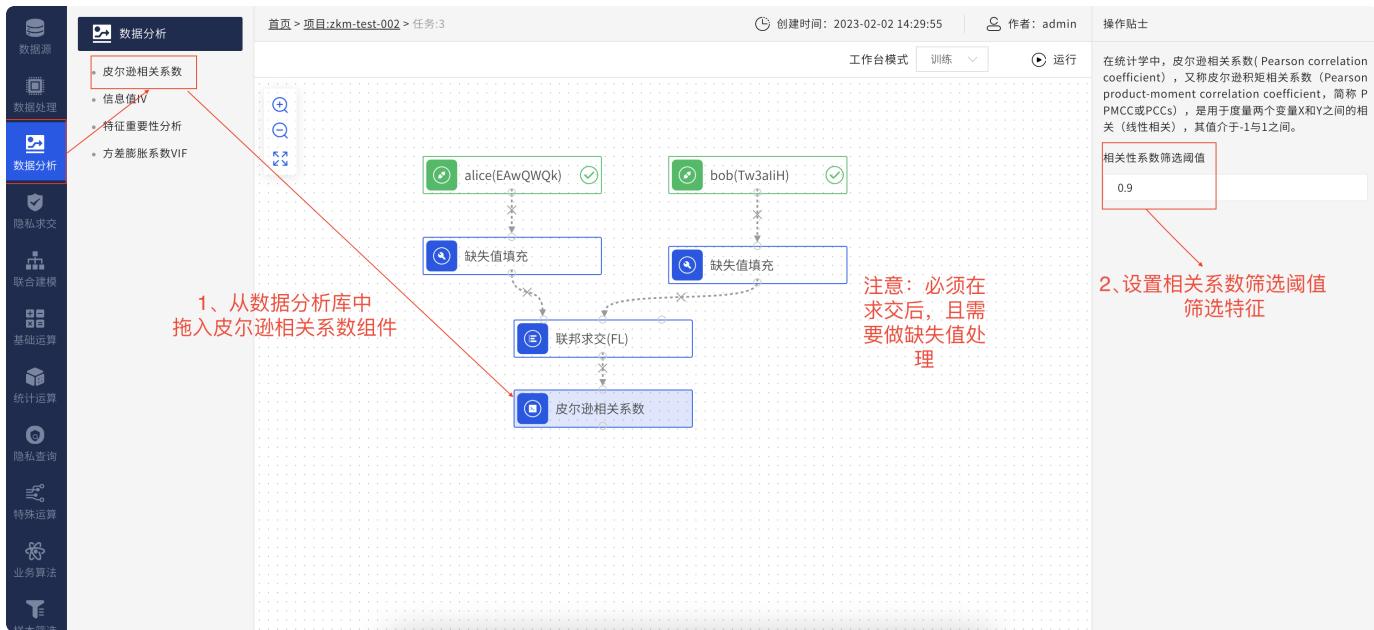


图100 皮尔逊相关系数筛选特征

c) 运行结果

- 在组件执行完成后，可通过右键查看更多详情



图101 皮尔逊相关系数组件运行结果右键

- 点击查看数据可查看特征的相关系数矩阵。



图102 皮尔逊相关系数组件运行结果右键查看数据

4.4.3.2. 信息值IV

a) 组件介绍

IV，即信息价值（Information Value），也称信息量。在机器学习的二分类问题中，IV值主要用来对输入变量进行编码和预测能力评估。特征变量IV值的大小即表示该变量预测能力的强弱，越大代表该特征预测能力越强，就越应该保留下。对于特征 X_i ，其IV值的计算步骤如下：

(1) 对特征X进行分桶，并计算每个桶的WOE值（WOE即变量权重，是对原始自变量的一种编码形式），对于第*i*组分桶，WOE的计算公式如下：

$$WOE_i = \ln \left(\frac{p_{yi}}{p_{ni}} \right) = \ln \left(\frac{y_i / y_T}{n_i / n_T} \right)$$

其中， p_{yi} 是这个组中坏样本的占总的坏样本的比例（风险模型中，对应的是违约客户，总之，指的是模型中预测变量取值为“是”或者说1的个体）占所有样本中所有响应客户的比例； p_{ni} 是这个组好样本的占总的好样本的比例； y_i 是这个组中坏样本的数量； n_i 是这个组中好样本的数量； y_T 是样本中所有坏样本的数量； n_T 是样本中所有好的数量。

(2) 计算每个桶的IV值，对于第*i*组分桶，IV的计算公式如下：

$$IV_i = (p_{yi} - p_{ni}) * WOE_i = (p_{yi} - p_{ni}) * \ln \left(\frac{p_{yi}}{p_{ni}} \right)$$

(3) 计算特征X的IV值，IV值在WOE的基础上保证了结果非负，根据特征X在各分桶上的IV值，得到整个特征X的IV值为：

$$IV = \sum_{i=1}^n IV_i = \sum_{i=1}^n (p_{yi} - p_{ni}) * WOE_i$$

用户可通过设置IV值筛选阈值来筛选特征，当特征的IV值小于该阈值时，就会被去掉。一般特征IV值的大小与预测能力有以下关系：

IV	预测能力
<0.03	无预测能力
0.03~0.09	低
0.1~0.29	中
0.3~0.49	高
>=0.5	极高

https://blog.csdn.net/LuYi_WeiLin

图103 IV值大小与预测能力关系图

b) 画布编辑

该组件需在求交后执行，且特征必须先进行缺失值处理。

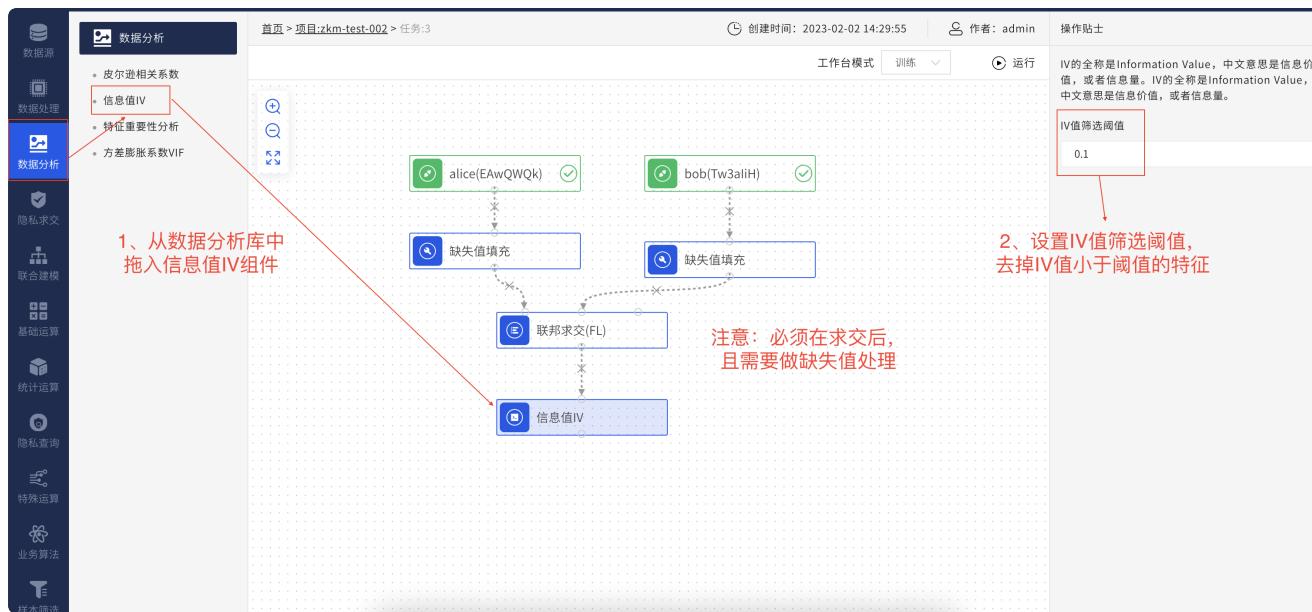


图104 信息值IV筛选特征

c) 运行结果

- 在组件执行完成后，可通过右键查看更多详情：如4.4.3.1数据分析-皮尔逊相关系数c）。
- 点击查看数据可查看所选特征的IV值。



图105 信息值IV运行结果右键查看数据

4.4.3.3. 特征重要性分析

a) 组件介绍

这里的特征重要性分析指的是GBDT树模型中的特征重要程度，即计算所有的非叶子节点在使用该特征分裂时加权不纯度的减少，减少得越多说明特征越重要。不纯度的减少实际上就是该节点此次分裂的收益，因此也可以这样理解，节点分裂时收益越大，该节点对应的特征的重要度越高。用户可通过设置**特征重要性排名**来保留最重要的几个特征。

b) 画布编辑

该组件需在求交后执行。

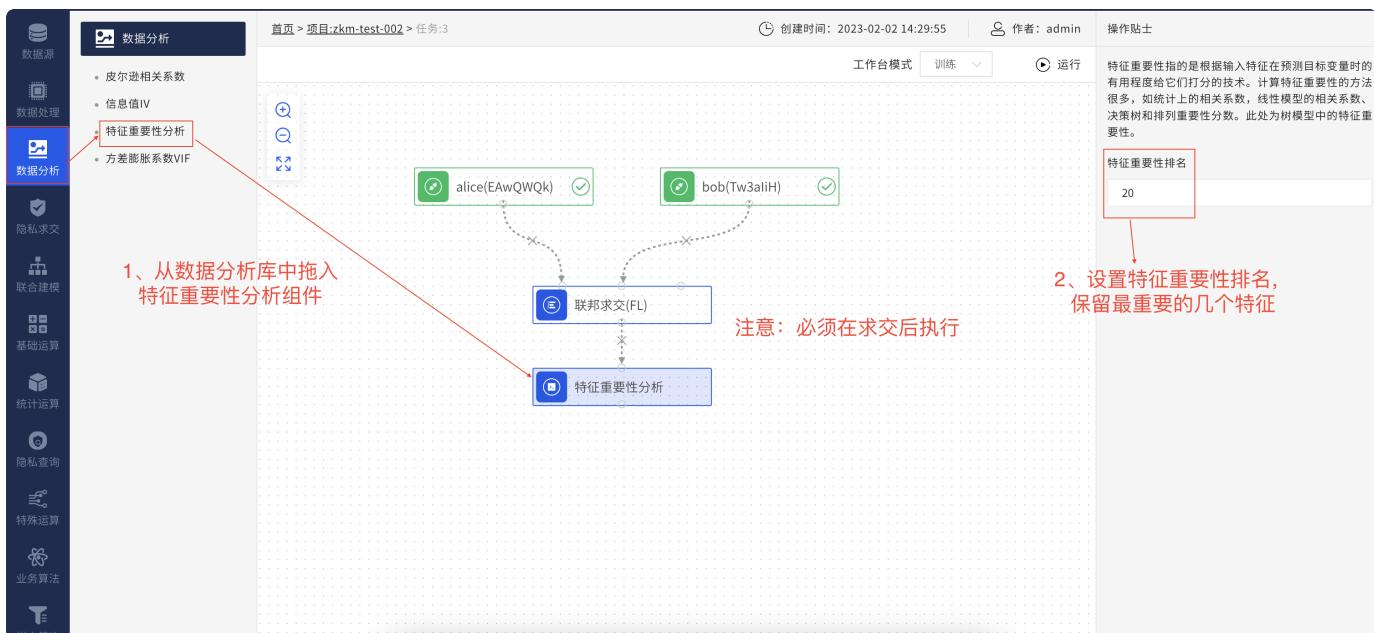


图106 特征重要性分析

c) 运行结果

- 在组件执行完成后，可通过右键查看更多详情：如4.4.3.1数据分析-皮尔逊相关系数c）。
- 点击查看数据可查看保留下来的最重要的几个特征及其贡献权重。

The screenshot shows a user interface for a machine learning model. On the left, there's a sidebar with various icons for data management, including '数据源' (Data Source), '数据处理' (Data Processing), '数据分析' (Data Analysis), '隐私求交' (Privacy Intersection), '联合建模' (Joint Modeling), '基础运算' (Basic Operation), '统计运算' (Statistical Operation), '隐私查询' (Privacy Query), '特殊运算' (Special Operation), '业务算法' (Business Algorithm), and '样本筛选' (Sample Selection). The main area displays a '任务' (Task) titled '数据' (Data) with a sub-section '特征重要性分析' (Feature Importance Analysis). This section contains a table:

特征名称	贡献权重
term	28.76
n0	10.86
n5	10.62

A tooltip or context menu is overlaid on the table, specifically highlighting the row for 'term'. The tooltip content is as follows:

特征重要性指的是根据输入特征在预测目标变量时的有用程度给它们打分的技术。计算特征重要性的方法很多，如统计上的相关系数，线性模型的相关系数、决策树和排列重要性分数。此处为树模型中的特征重要性。

特征重要性排名
3

图107 特征重要性分析组件运行结果右键查看数据

4.4.3.4. 方差膨胀系数VIF

a) 组件介绍

方差膨胀因子VIF指的是解释变量之间存在多重共线性时的方差与不存在多重共线性时的方差之比，可以反映多重共线性导致的方差的增加程度。特征VIF值越大，代表该特征多重共线性的影响越严重，参数估计的结果不再具有有效性。特征 X_i 的方差膨胀系数VIF的计算公式为：

$$VIF = \frac{1}{1 - R_i^2},$$

其中 R_i 是特征 X_i 与其他特征 X_j 的复相关系数，所谓复相关系数即可决系数 R^2 的算术平方根，也即拟合优度的算术平方根。不过这个可决系数 R^2 是指用 X_i 作为因变量，对其特征作为自变量，搭建一个新的回归模型后得到的可决系数。

用户可通过设置VIF值筛选阈值来筛选特征，当特征的VIF值大于该阈值时就会被删除，常见设置值为10。

b) 画布编辑

该组件需在求交后执行，且特征必须先进行缺失值处理。

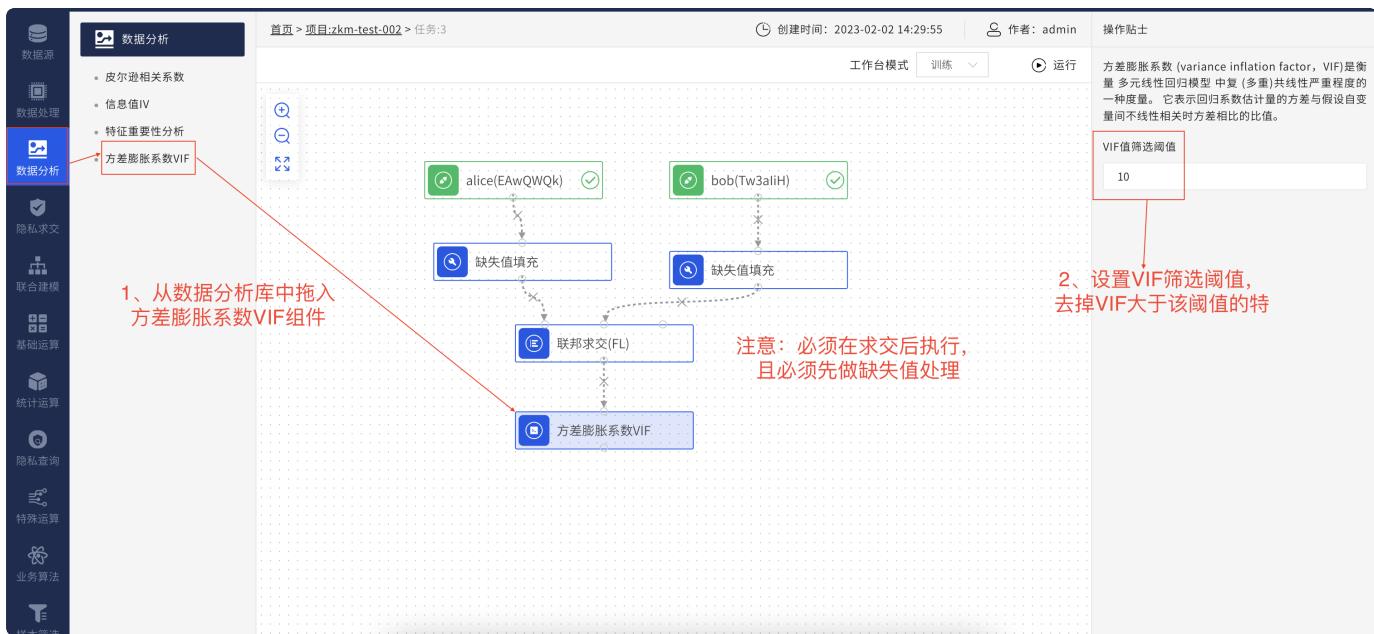


图108 方差膨胀系数VIF筛选特征

c) 运行结果

- 在组件执行完成后，可通过右键查看更多详情：如4.4.3.1数据分析-皮尔逊相关系数c）。
- 点击查看数据可查看选中特征的VIF值。

This screenshot shows the results of the VIF component execution. The left sidebar shows data sources 'alice' and 'bob'. The main workspace displays a '数据' (Data) window with a '特征重要性分析' (Feature Importance Analysis) table:

特征名称	VIF
n0	1.003774
n1	1.04234
loanAmnt	1.21586
term	1.170961

A note on the right side states: '方差膨胀系数 (variance inflation factor, VIF) 是衡量多元线性回归模型 中复 (多重)共线性严重程度的一种度量。它表示回归系数估计量的方差与假设自变量间不线性相关时方差相比的比值。' (VIF is a measure of multicollinearity in multiple linear regression models. It represents the ratio of the variance of the estimated regression coefficients to the variance when the independent variables are not linearly related.) and a 'VIF值筛选阈值' (VIF threshold filtering) input field set to '10'.

图109 方差膨胀系数VIF组件运行结果右键查看数据

4.4.4. 隐私求交

通常是涉及多参与方的任务执行的第一步。用于对两方或两方以上的样本数据，根据ID字段计算各参与方样本数据的ID交集，同时不泄露交集以外的其他ID。金智塔隐私计算平台提供了两种“隐私求交”组件——联邦求交 (outsourced) 和多方安全求交 (MPC)。

4.4.4.1. 联邦求交

a) 组件说明

除了提供数据的参与方，计算过程还会涉及一个辅助计算的参与方。任务中的各参与方协商密钥后，提供数据的参与方将自有ID加密，并统一发送给辅助计算的参与方进行样本对齐。随后，辅助计算的参与方会将对齐后的样本ID回传给所有提供数据的参与方。该组件运行效率高，安全性较高，但通信量较大。当网络状况良好，任务数据集较大，且信任辅助计算的参与方环境（自己提供）时，建议使用该组件。

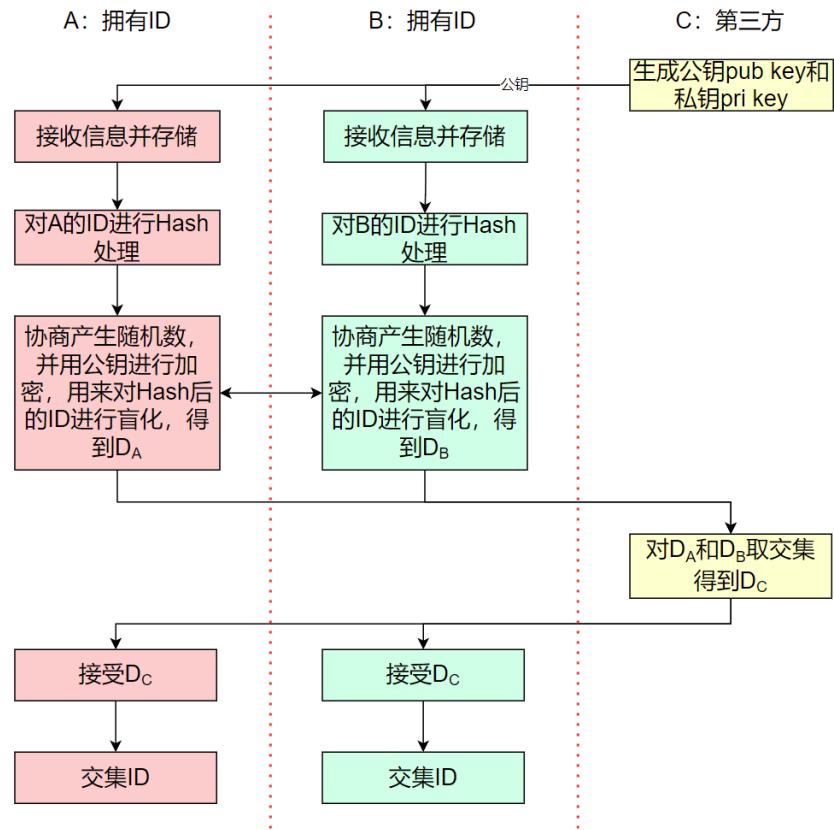


图110 联邦求交

b) 画布编辑

从“隐私求交”组件库中选择“联邦求交 (FL)”或者“多方安全求交 (DH)”组件并拖至画布中，左键长按“数据源组件的输出端”，拉出引线至“隐私求交组件的输入端”。最终效果如下图所示，点击“隐私求交”组件框，可以对求交算法的参数进行配置，默认是最高安全级别的配置。

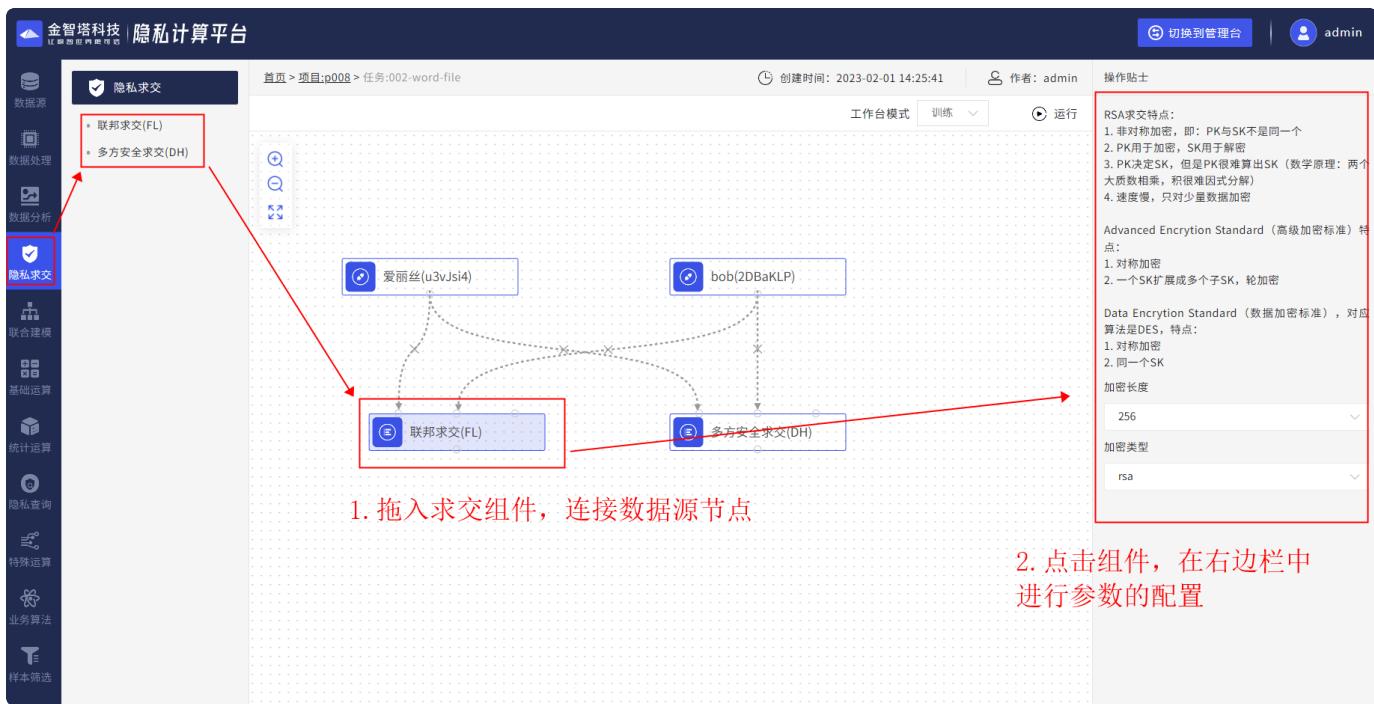


图111 隐私求交操作说明

c) 运行结果

执行成功后，右键“隐私求交”组件框，点击“查看ID列”获取多方对齐后的ID集合。

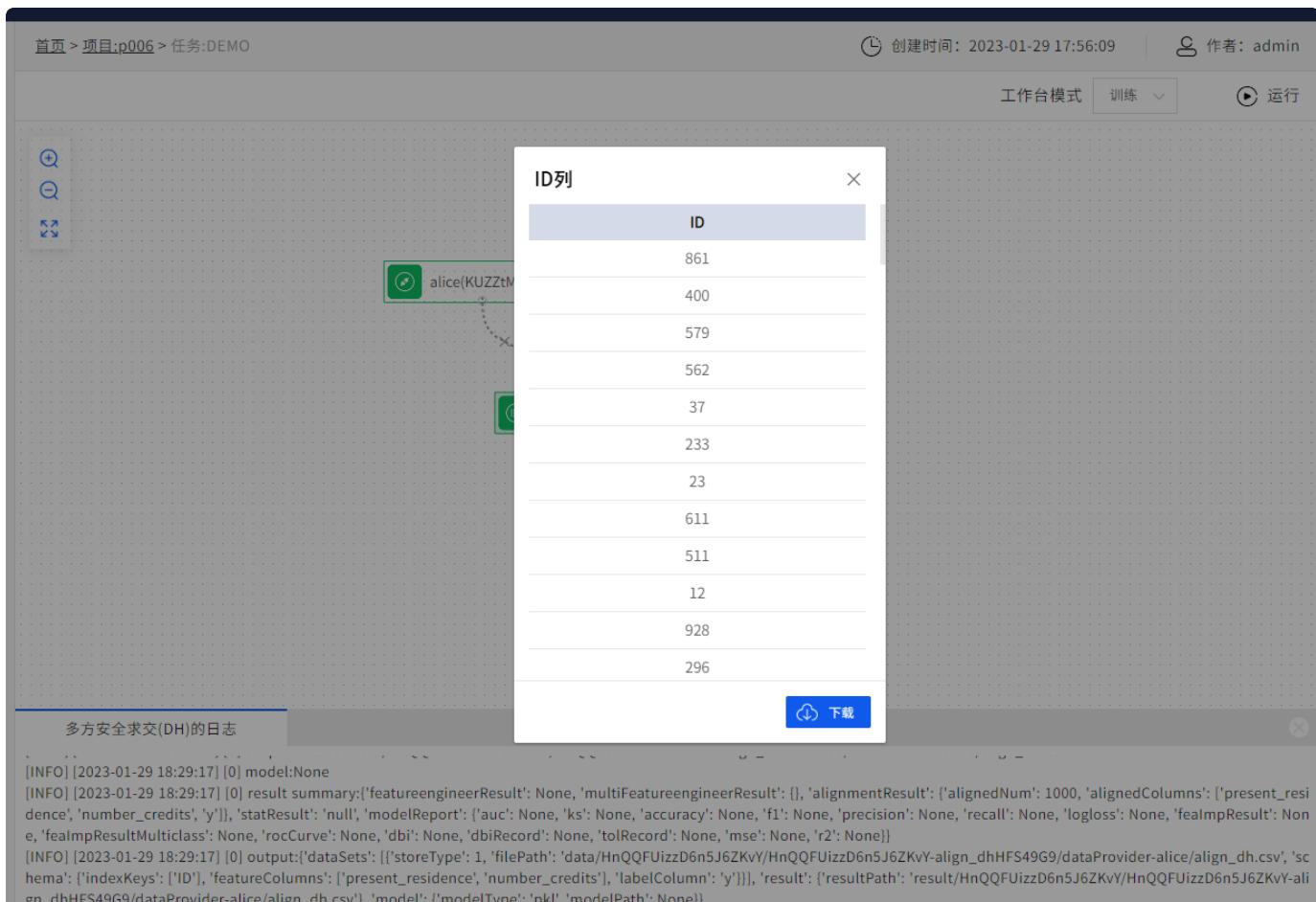


图112 隐私计算结果

右键菜单说明：

- a. 查看数据：获取任务明细。
- b. 查看ID列：获取任务对齐的ID集合。
- c. 数据审计：获取各方对齐样本的统计分析面板。

4.4.4.2. 多方安全求交

a) 组件说明

计算过程仅涉及提供数据的各参与方。该组件底层的运算和数据交互只会发生在提供数据的参与方之间，因此执行环境相较于“联邦求交”会更规范。但算法涉及多次加密，因此运行效率相对较低。建议在执行数据量较小，或对安全性要求较高的任务时使用。

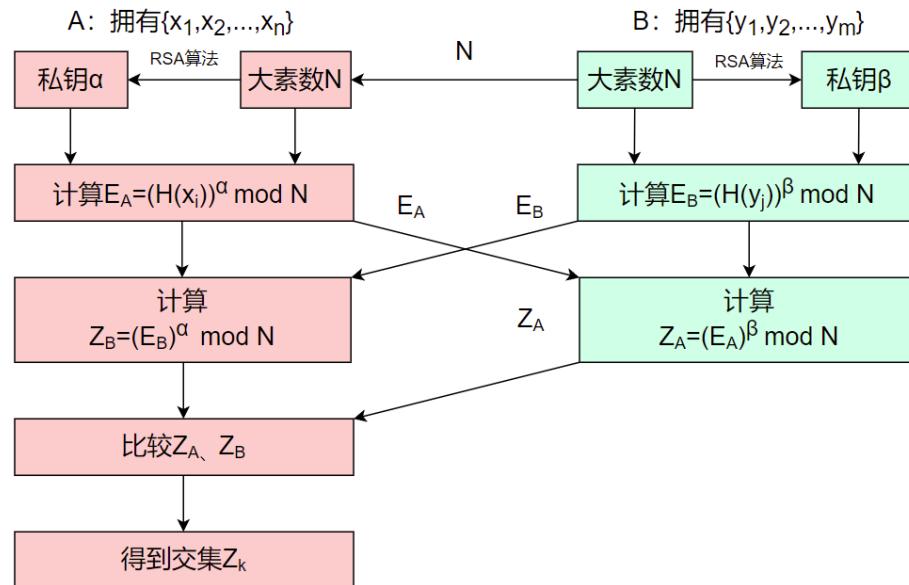


图113 多方安全求交

b) 画布编辑

同4.4.4.1. 联邦求交中的b)

c) 运行结果

同4.4.4.1. 联邦求交中的c)

4.4.5 联合建模

联合建模包含联邦学习和多方安全计算两大类。其中联邦学习支持逻辑回归、Xgboost、K-means，多方安全计算支持逻辑回归。

4.4.5.1. 联邦学习-逻辑回归

a) 组件说明

逻辑回归 (Logistic Regression) 又称logistic回归分析，是一种广义的线性回归分析模型，用于解决二分类 (0 or 1) 问题的机器学习方法，用于估计某种事物的可能性。

逻辑回归与多重线性回归分析有很多相同之处。它们的模型形式基本上相同，都具有 $w'x + b$ ，其中w和b是待求参数，其区别在于他们的因变量不同，多重线性回归直接将 $w'x + b$ 作为因变量，即 $y = w'x + b$ ，而logistic回归则通过函数L将 $w'x + b$ 对应一个隐状态p， $p = L(w'x + b)$ ，然后根据p与1-p的大小决定因变量的值。

联邦学习的逻辑回归，在模型聚合使用了安全聚合算法：各参与方通过加入随机数，汇总后的随机数能够抵消，使得第三方只能得到总模型，无法得知某个参与方的具体模型。

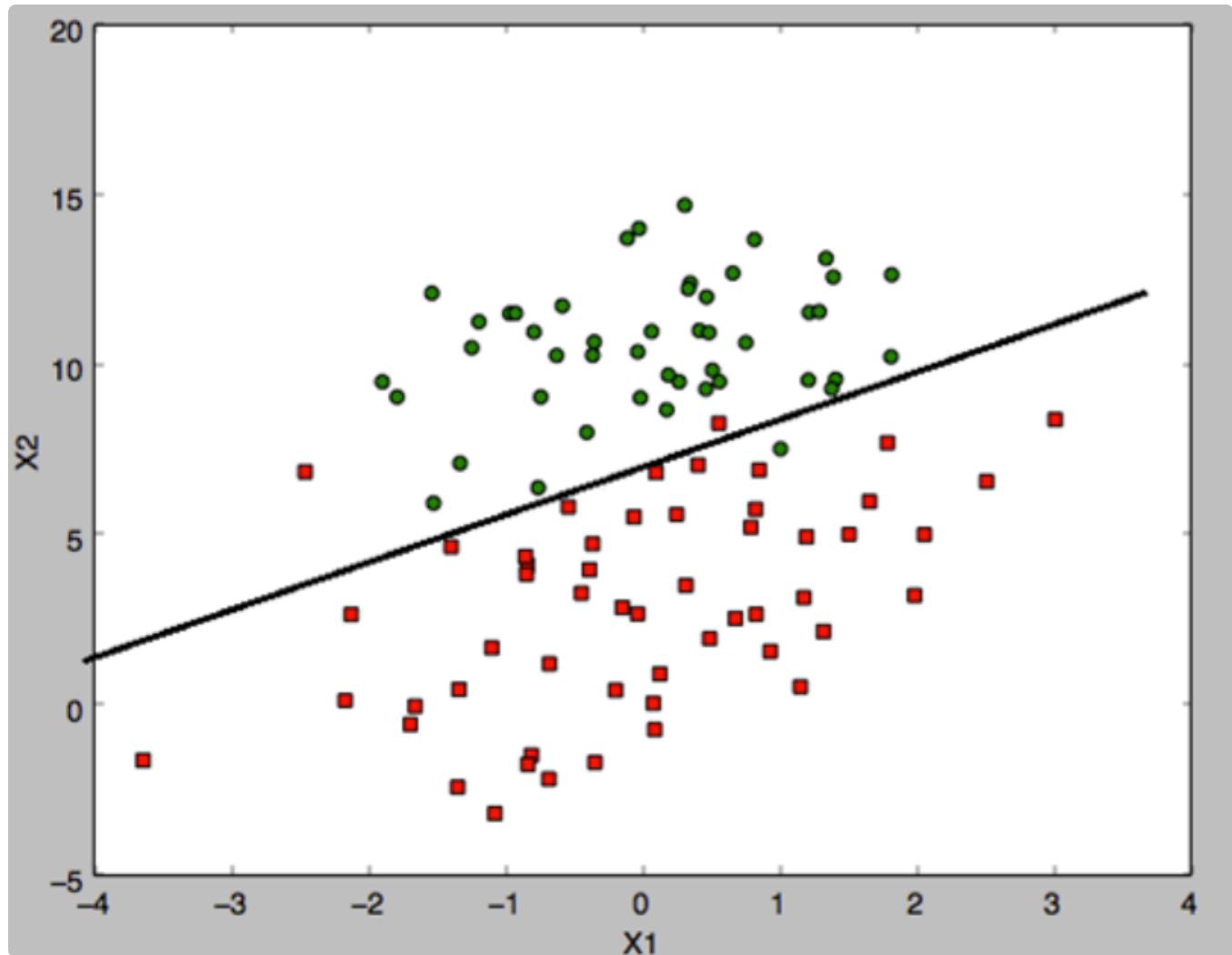


图114 逻辑回归

b) 画布编辑

该组件需在求交后执行

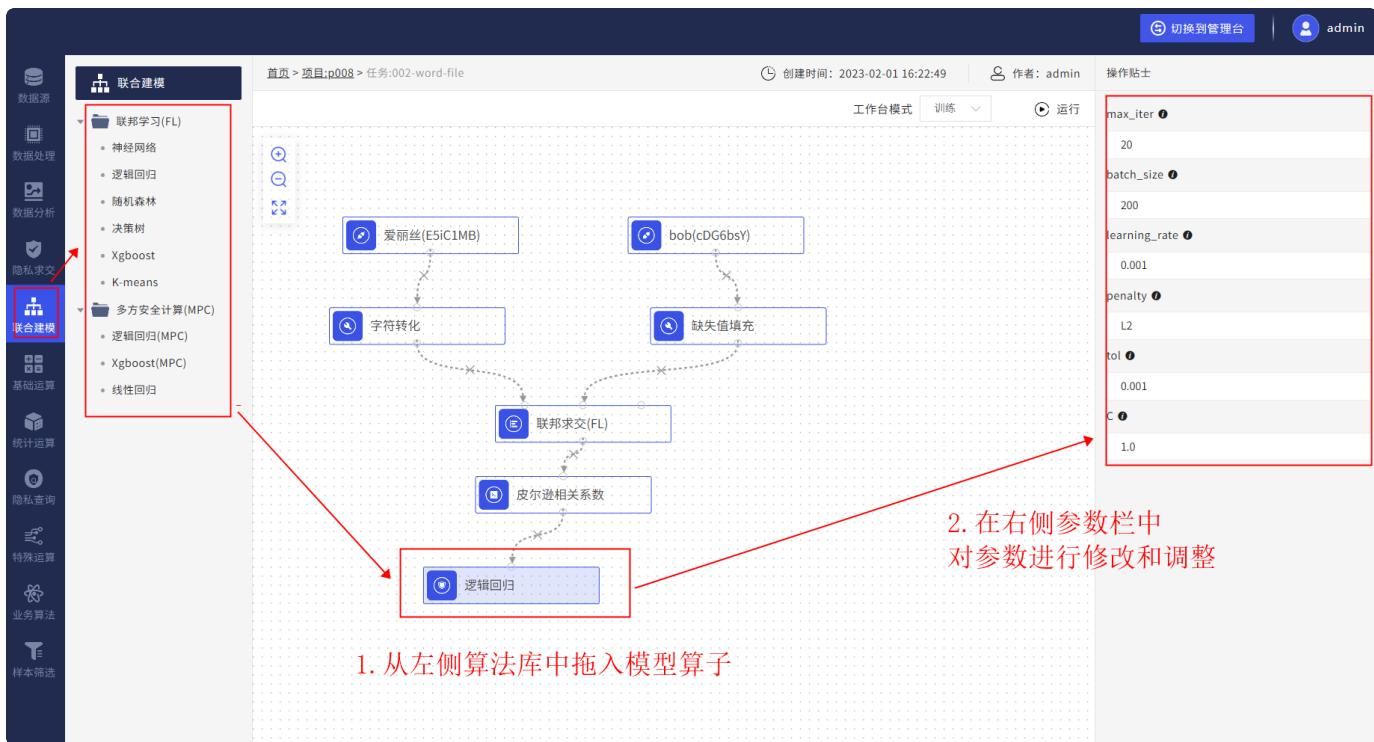


图115 逻辑回归操作说明

c) 运行结果

用户在完成模型训练任务后，可通过点击模型算子，在画布右边栏中点击**查看报告**，可以查看参数配置、特征选择、效果验证、贡献评估等内容，如下图所示。



图116 逻辑回归模型报告

模型报告

X

参数配置

max_iter: 20 batch_size: 200 learning_rate: 0.001 penalty: L2 tol: 0.001 C: 1.0

特征选择

爱丽丝(E5IC1MB)	
特征中文	
present_residence	property
age	other_installment
housing	number_credits
job	number_maintenance
telephone	foreign_worker
education	investment_interest
channel_source	is_local
overdue_ratio	
特征英文	
present_residence	property
age	other_installment
housing	number_credits
job	number_maintenance
telephone	foreign_worker
education	investment_interest
channel_source	is_local
overdue_ratio	

bob(cDG6bsY)	
特征中文	
account_age	duration_month
credit_history	purpose
credit_amount	savings_account
present_employment	income
personal_status	guarantors
特征英文	
account_age	duration_month
credit_history	purpose
credit_amount	savings_account
present_employment	income
personal_status	guarantors

标签选择

爱丽丝(E5IC1MB)	
特征中文	y
特征英文	y

效果验证



贡献评估

节点名称	特征名称	特征重要性
bob	credit_history	-0.343945
bob	credit_amount	0.194749
bob	present_employment	-0.197173
bob	income	0.121539
bob	purpose	-0.181746
bob	account_age	-0.072872
bob	guarantors	-0.045588
bob	duration_month	0.016968
bob	personal_status	0.113831

图117 模型报告详情

d) 模型评估

当用户完成模型训练后，可在工作台画布的右上角，将工作台模式切换为评估，可以使用评估数据集对训练好的模型进行效果验证，如下图所示。

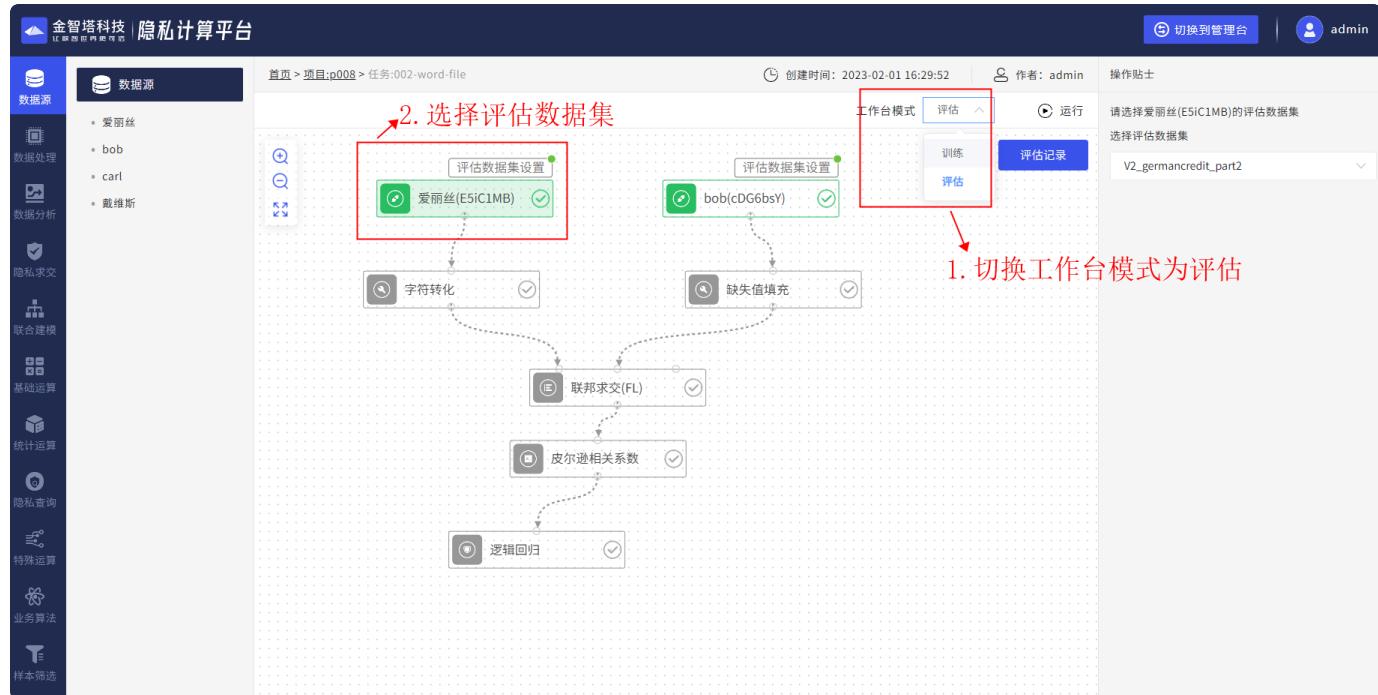


图118 模型评估

评估完成后，可点击评估记录，查看评估的模型指标。

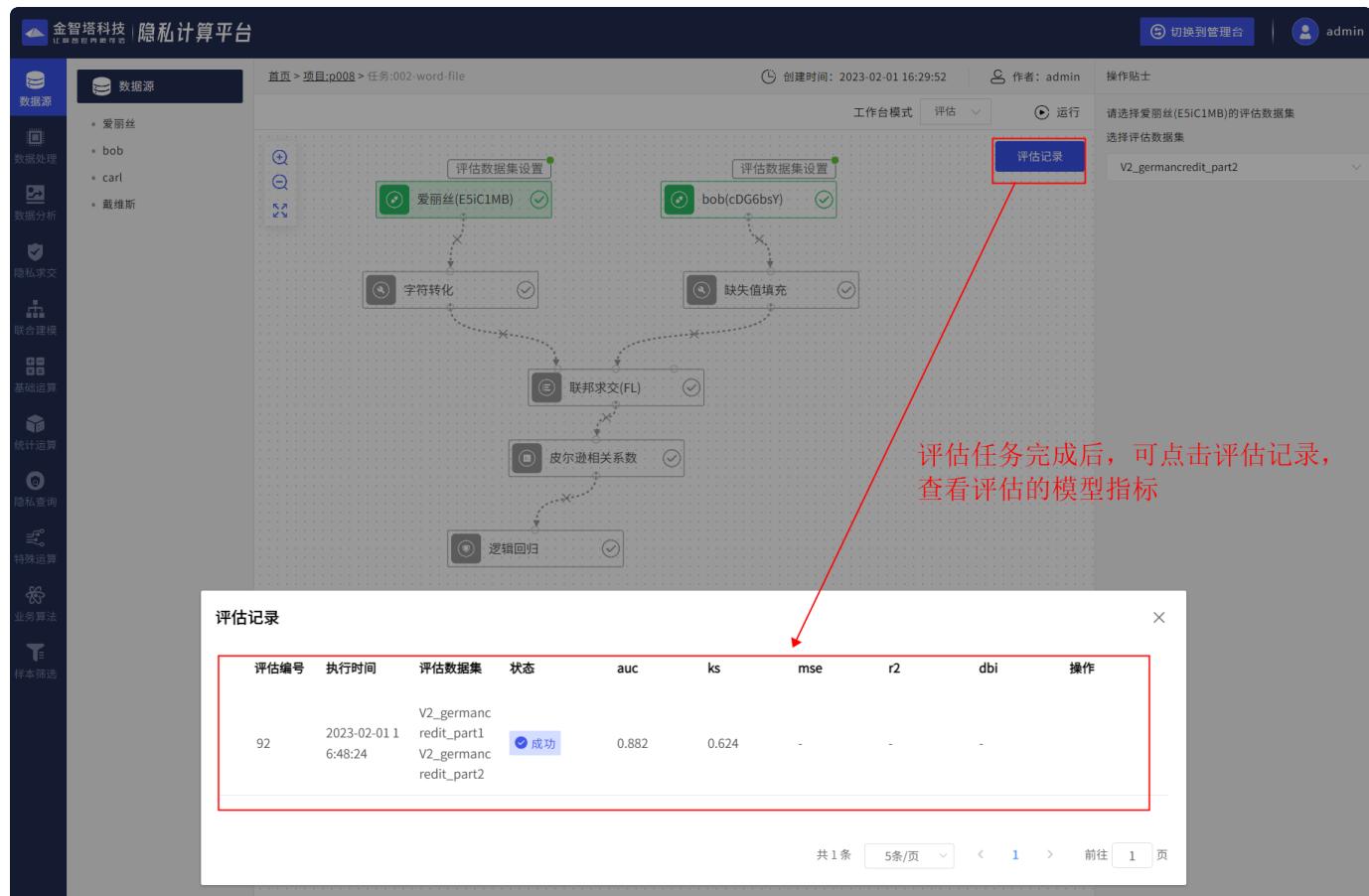


图119 模型评估记录

4.4.5.2. 联邦学习-Xgboost

a) 组件说明

梯度提升决策树 (Gradient Boosting Decision Tree, GBDT) 是一种基于boosting集成思想的加法模型，训练时采用前向分布算法进行贪婪的学习，每次迭代都学习一棵CART树来拟合之前 $t-1$ 棵树的预测结果与训练样本真实值的残差。XGBoost的基本思想和GBDT相同，但是做了一些优化，比如二阶导数使损失函数更精准；正则项避免树过拟合；Block存储可以并行计算等。

XGBoost的基本思想和GBDT相同，但是做了一些优化，比如二阶导数使损失函数更精准；正则项避免树过拟合；Block存储可以并行计算等。

联邦学习XGBoost是一种允许多方参与的纵向联邦学习算法，基本原理跟XGBoost类似，在此基础上引入联邦学习要考虑的隐私保护问题，是一种端到端的联邦环境中的梯度提升算法。

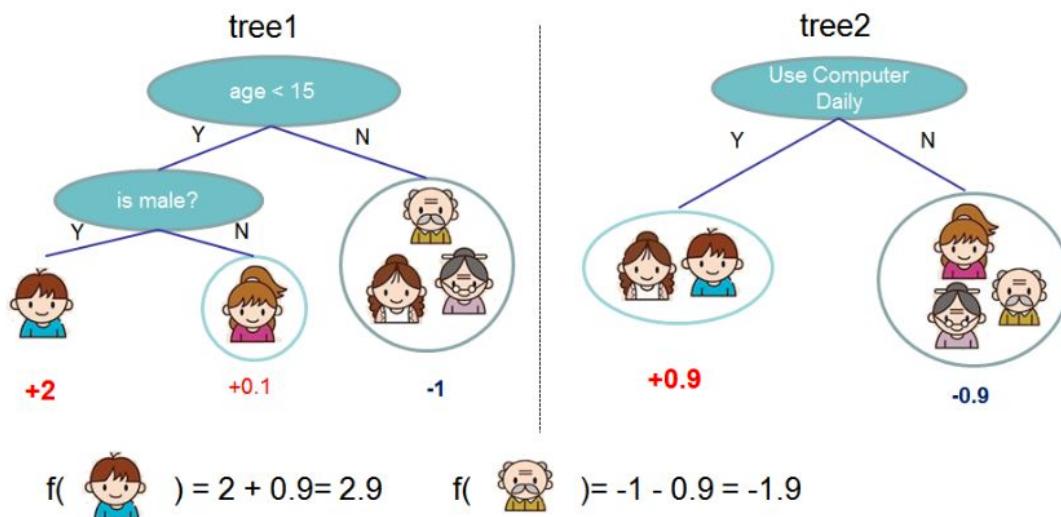


图120 梯度提升树

b) 画布编辑

该组件需在求交后执行

The screenshot shows the platform's interface for building federated learning models. On the left sidebar, under '联合建模' (Joint Modeling), there is a section for '联邦学习(FL)' which includes 'Xgboost'. A red arrow points from this section to a central workspace where an 'Xgboost' component is being configured. Another red arrow points to the right panel, which displays the model's parameters. The parameters listed are:

- n_estimators: 6
- max_depth: 5
- min_samples_split: 50
- min_samples_leaf: 10
- random_state: 0
- reg_lambda: 0.1
- gamma: 0.1
- learning_rate: 0.1
- scale_pos_weight: 1.0
- max_delta_step: 0.0
- subsample: 1.0
- reg_alpha: 0.0
- colsample_bytree: 1.0

图121 Xgboost模型使用说明

c) 运行结果

同4.4.5.1. 联邦学习–逻辑回归中的c)

d) 模型评估

同4.4.5.1. 联邦学习–逻辑回归中的d)

4.4.5.3. 联邦学习–K-means

a) 组件说明

K均值聚类算法 (k-means clustering algorithm) 是一种迭代求解的聚类分析算法，其步骤是，预将数据分为K组，则随机选取K个对象作为初始的聚类中心，然后计算每个对象与各个种子聚类中心之间的距离，把每个对象分配给距离它最近的聚类中心。聚类中心以及分配给它们的对象就代表一个聚类。每分配一个样本，聚类的聚类中心会根据聚类中现有的对象被重新计算。这个过程将不断重复直到满足某个终止条件。终止条件可以是没有（或最小数目）对象被重新分配给不同的聚类，没有（或最小数目）聚类中心再发生变化，误差平方和局部最小。

基于联邦学习技术实现的K-means，在每次迭代过程中，每个参与者利用本地数据训练接收到的模型，并将训练梯度发送给中心服务器。中心服务器聚合接收到的梯度来更新全局模型。

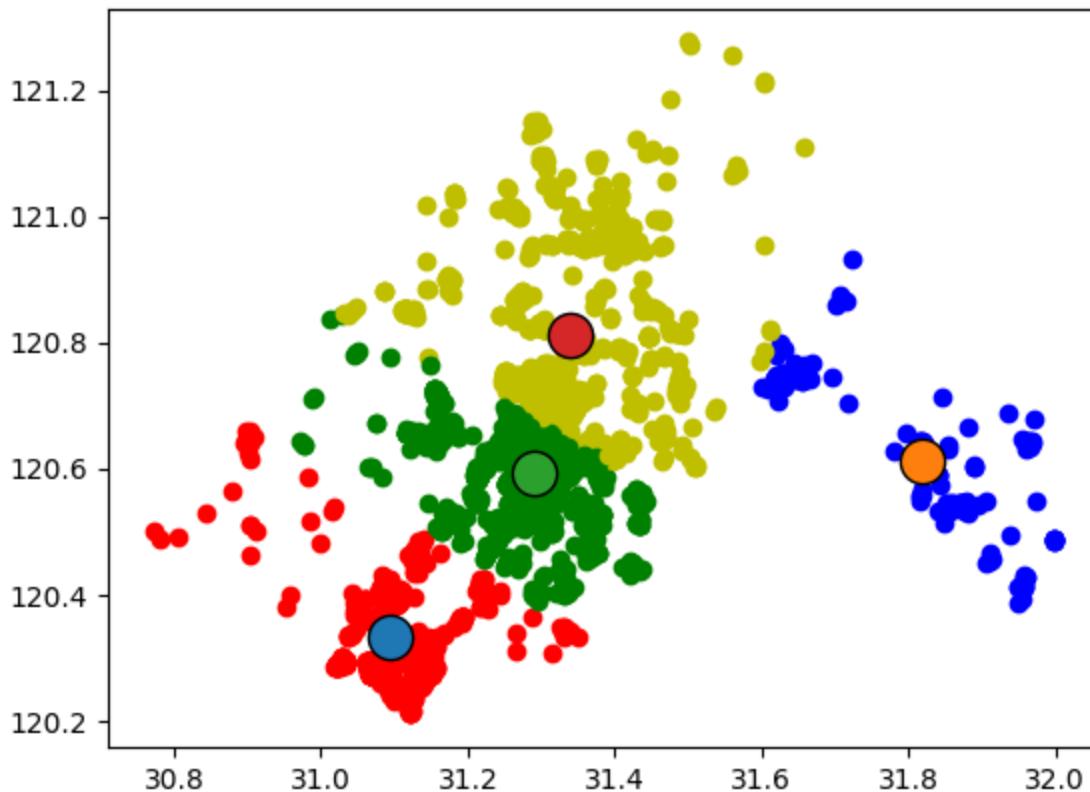


图122 K-means算法

b) 画布编辑

该组件需在求交后执行，且数据源模块中无需指定建模标签字段。

1. 从算法库中拖入K-means

2. 调整模型参数

图123 K-means使用说明

c) 运行结果

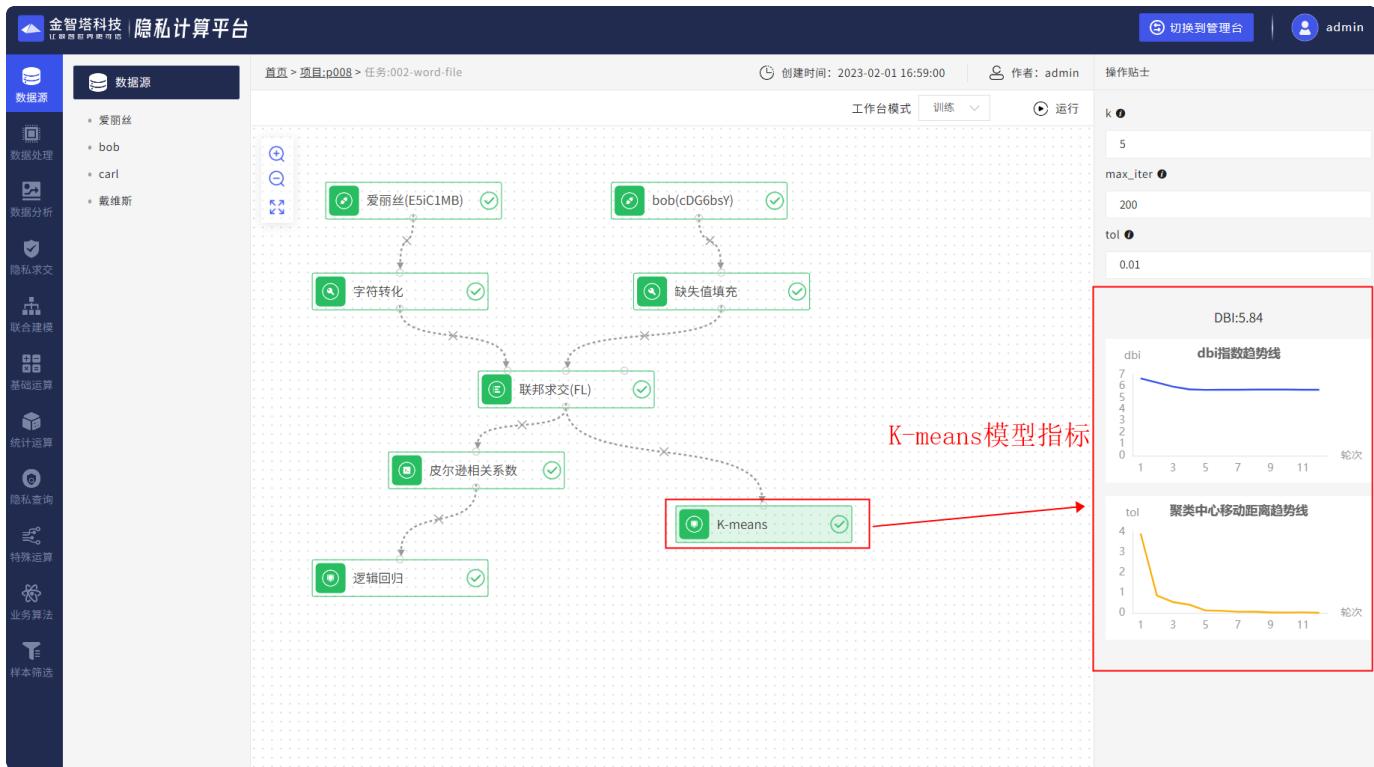


图124 K-means模型报告

d) 模型评估

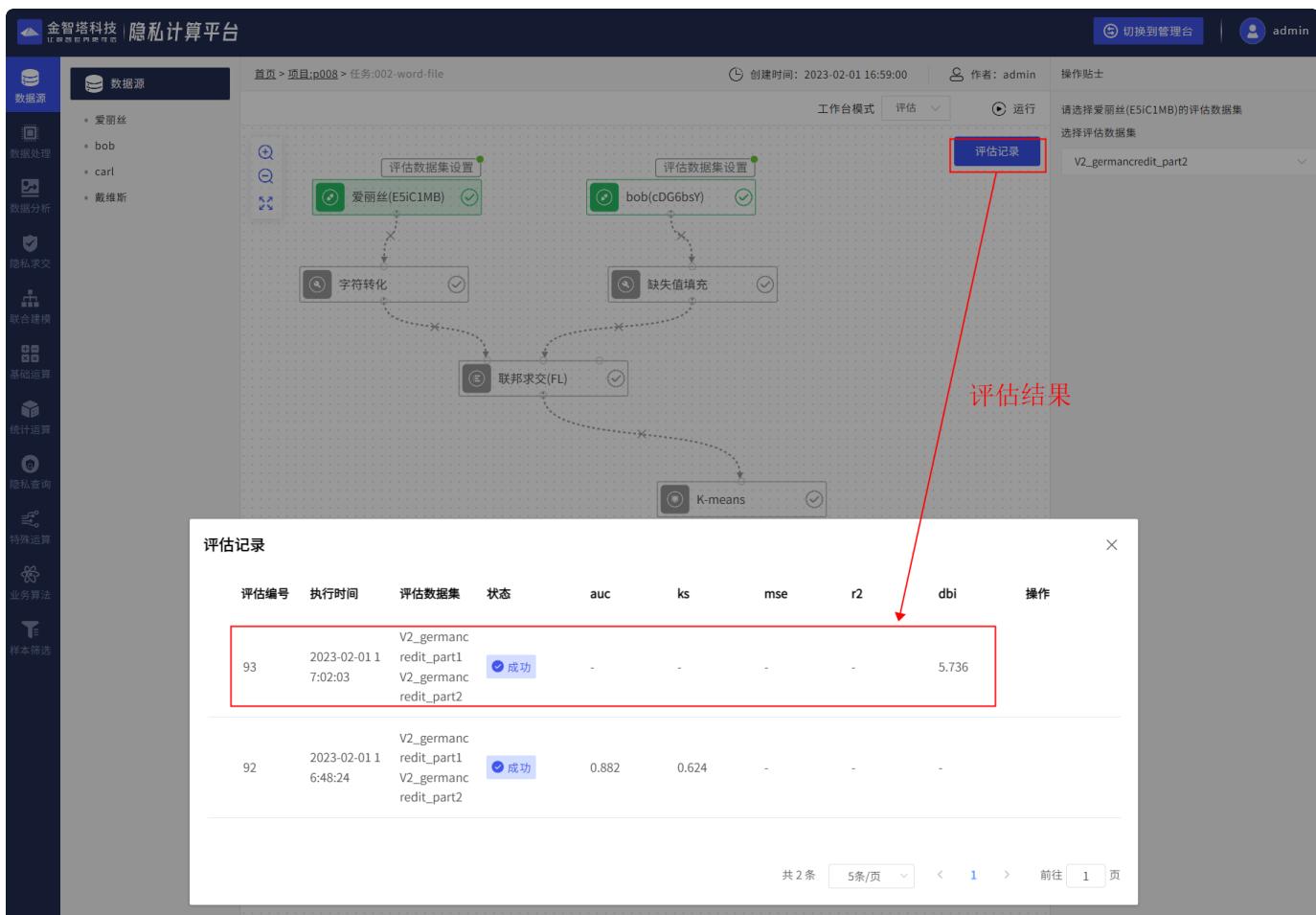


图125 K-means模型评估

4.4.5.4. 多方安全计算-逻辑回归

a) 组件说明

逻辑回归 (Logistic Regression) 又称logistic回归分析，是一种广义的线性回归分析模型，用于解决二分类 (0 or 1) 问题的机器学习方法，用于估计某种事物的可能性。

逻辑回归与多重线性回归分析有很多相同之处。它们的模型形式基本上相同，都具有 $w'x + b$ ，其中w和b是待求参数，其区别在于他们的因变量不同，多重线性回归直接将 $w'x + b$ 作为因变量，即 $y = w'x + b$ ，而logistic回归则通过函数L将 $w'x + b$ 对应一个隐状态p， $p = L(w'x + b)$ ，然后根据p与1-p的大小决定因变量的值。

基于多方安全计算技术实现的逻辑回归，利用秘密分享技术，实现模型计算过程中，没有明文泄露。

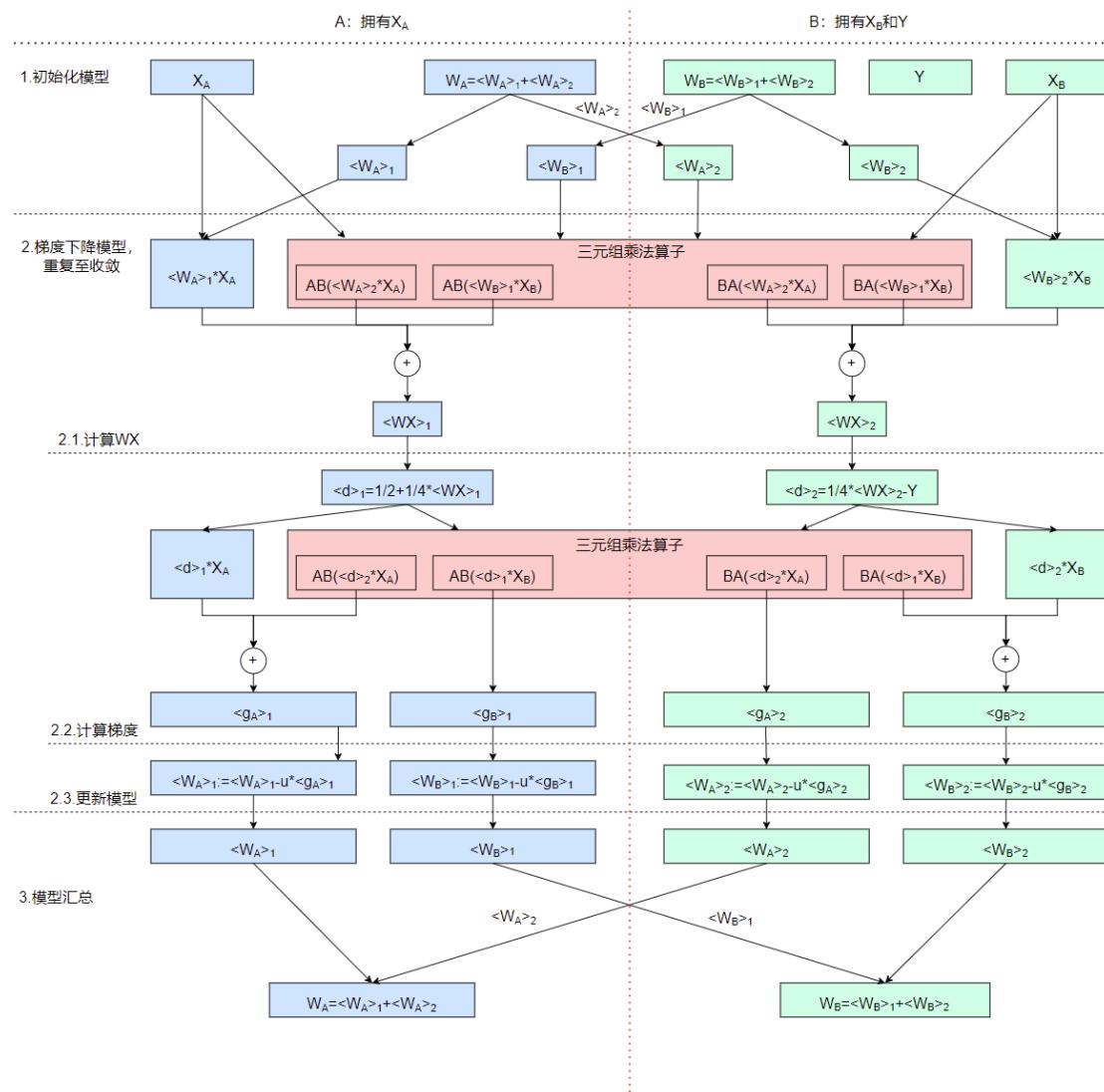


图126 MPC逻辑回归

b) 画布编辑

同4.4.5.1. 联邦学习-逻辑回归中的b)

c) 运行结果

同4.4.5.1. 联邦学习-逻辑回归中的c)

d) 模型评估

同4.4.5.1. 联邦学习-逻辑回归中的d)

4.4.6. 基础运算

用户可拖入基础运算中的算子，对多方数据进行基础计算。其中不同组件的支持的参与节点数不同，详情可见下表所示。执行结束后，用户可通过右键点击组件，点击运算结果，可查看组件计算的结果。

表1 基础运算组件规则

组件名称	支持参与方（节点）数量	是否支持多列同时计算
隐私比较	2	支持
多方安全加法	≥ 2	支持
多方安全减法	2	支持
多方安全乘法	2,3	支持
多方安全除法	2	支持
多方安全比较	2	支持
最大值	2,3	支持
方差	2,3	支持
中位数	2,3	支持
均值	≥ 2	支持

4.4.6.1. 安全四则运算

a) 组件说明

安全四则运算是指两方及两方以上节点的联合加、减、乘、除运算。多方数据首先使用求交算子进行样本对齐，且需要两方以上参与计算。这里的联合的四则计算是指样本特征维度进行四则运算。

b) 画布编辑

选择参与计算的数据源，然后配置各个数据源参与计算的数据特征和样本对齐ID字段，并添加求交算子。

从组件库基础运算中选择所需的四则运算算子，将算子拖入左侧画布，连接到求交算子下游。

点击算子，点击右侧的添加运算，选择每个数据源参与指标计算的特征。

完成配置后点击运行，运行结束后，右键点击运算结果可以查看下载结果。



图127 基础四则运算任务

c) 运行结果

运算结果

id	features_0+features_20+features_40	features_1+features_27+features_41
ID_0	1.278193872627895	-0.3615905982503648
ID_1	1.7136076586901596	0.10560990768700917
ID_2	-0.2521139894742588	-0.8587576835439312
ID_3	-2.2842242338275214	0.602208605363401
ID_4	-0.8532601713807341	-1.9295576491033692
ID_5	2.1692420580645577	1.8246479815863645
ID_6	-2.4373789695868466	-1.371338141065753
ID_7	-1.0266565266631518	-1.327110210415689
ID_8	-2.0083782215934813	1.37121193818885

下载

图128 多方安全加法-运算结果

4.4.6.2. 安全比较

a) 组件说明

安全比较是指多方数据源进行样本对齐后，对样本来自不同方的特征进行大小对比。安全比较算子仅支持两方比较。

b) 画布编辑

同4.4.5.1. b)

c) 运行结果

ID	account_age_cmp_duration_month	X
a229	小于	
716	小于	
a244	小于	
538	小于	
912	小于	
241	小于	
a691	小于	
a646	小于	
a10	小于	
571	小于	
860	小于	
...	...	

 下载

图129 安全比较-运算结果

4.4.7. 统计运算

统计运算：包含最大值、中位数、均值和方差。

注：以上组件均为隐私计算组件，用于统计两方数据的统计分析指标，因此需要放到“隐私求交”组件的下游。如需查看单方数据中某一列的统计分析指标，可以在“数据审计”的页面中查看。

4.4.7.1. 最大值

a) 组件说明

用于计算各参与方选定列合并后的列值集合的最大值。

b) 画布编辑

在“隐私求交”算子的下游拖入“最大值”组件，并在右侧面板配置参与“最大值”运算的各参与方的列。



图130 最大值组件说明

c) 运行结果

右键最大值组件框，单击“查看数据”，获取最大值。

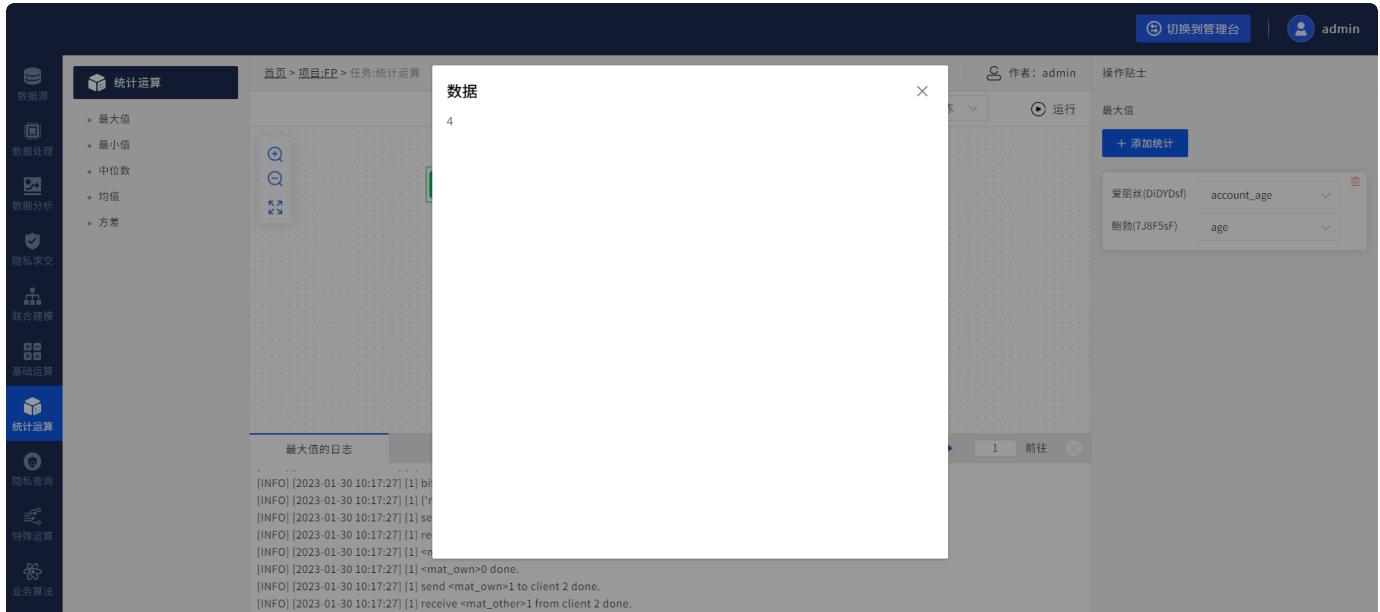


图131 最大值组件结果

4.4.7.2. 中位数

a) 组件说明

用于计算各参与方选定列合并后的列值集合排序后，排在中间的列值。

b) 画布编辑

同4.4.7.1. 最大值中的b)

c) 运行结果

同4.4.7.1. 最大值中的c)

4.4.7.3. 均值

a) 组件说明

用于计算各参与方选定列合并后的列值集合的均值。

b) 画布编辑

同4.4.7.1. 最大值中的b)

c) 运行结果

同4.4.7.1. 最大值中的c)

4.4.7.4. 方差

a) 组件说明

用于计算各参与方选定列合并后的列值集合的方差。

b) 画布编辑

同4.4.7.1. 最大值中的b)

c) 运行结果

同4.4.7.1. 最大值中的c)

4.4.8. 隐私查询

4.4.8.1. 匿踪查询

a) 组件说明

隐私查询/隐私信息检索(Private Information Retrieval – PIR)技术是解决保护用户查询隐私的方案。主要目的是，保证查询用户在向服务器上的数据库提交查询请求，在用户查询隐私信息不被泄漏的条件下完成查询，即在过程中服务器不知道用户具体查询信息及检索出的数据项。

假定数据库是一个由n位二进制数组成的字符串S。当用户对字符串S中的第i位查询字符Si进行查询时，如果直接进行查询，肯定会将Si值的相关信息泄露，造成隐私泄露的严重后果。为了保护数据隐私，用户不能直接发起查询，而在查询之前，先使用加密机制对查询i进行加密得到 $E(i)$ ，而后将加密的 $E(i)$ 发送给位置服务器进行查询。服务器收到查询请求 $E(i)$ 后，进行查询数据库操作，并将查询得到的结果 $q(S, E(i))$ 返回给用户。当用户收到查询结果 $q(S, E(i))$ 后，应用解密操作进行解密，得到最终的查询结果。

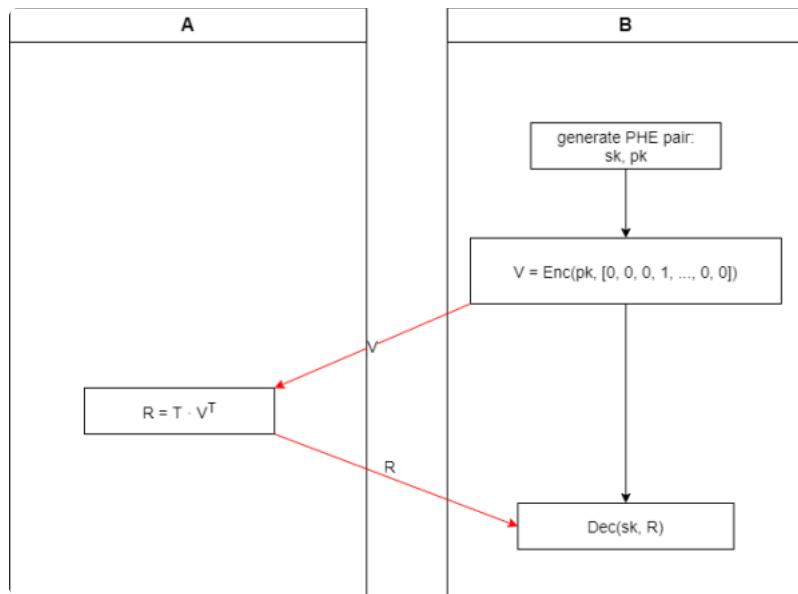


图 132 匿踪查询

b) 画布编辑

用户可在组件库中选择隐私查询进行匿踪查询（PIR）的执行。查询方必须是本方节点，查询方需要在数据源中配置样本ID字段为查询ID，被查询方需要指定样本ID为被查询匹配ID以及可以被查询的特征变量。匿迹查询算子可以通过点击后在右侧配置，查询方式根据查询安全和效率的结合方案不同，分为高效率查询（查询速度快，安全性相对较弱）和高隐匿查询（查询速度慢但安全性较高），主查节点为查节点，即为本方节点。结果可以右击组件查看。

Screenshot of the Privacy Computing Platform interface, showing the configuration of a PIR component:

- 左侧菜单栏:** 包含数据源、数据处理、数据分析、隐私求交、联合建模、基础运算、统计运算、隐私查询、特殊运算、业务算法等模块。
- 中心工作区:** 显示了两个节点（alice 和 bob）与一个“匿踪查询(PIR)”组件的连接关系。
- 右侧配置栏:**
 - 操作贴士: 隐私检索(PIR)是多方计算协议的重要组成部分，基于同态加密技术，使得查询的用户与数据拥有者在各自的私有信息不泄露的情况下完成安全查询操作。
 - 加密长度: 2048
 - 加密类型: paillier
 - 查询方式: 高效率查询
 - 主查节点: alice(mwZAk3R)
- 注释:** “指定查询方” (Specify Query Node) 指向配置栏中的“主查节点”输入框。

图133 匿踪查询组件说明

c) 运行结果

检索结果 X

ID	account_age
0	1.0
1	2.0
2	4.0
3	1.0
4	1.0
5	4.0
6	4.0
7	2.0
8	4.0
9	2.0
10	2.0

下载

图134 匿踪查询运行结果

4.4.9. 特殊运算

4.4.9.1. 偏离度计算

a) 组件说明

基于多方安全技术实现的两个数值之间的偏差计算（该组件需在求交后执行）。

$$\text{偏离度} = \frac{|A - X|}{X}, A\text{为目标数据, } X\text{为实际数据}$$

b) 画布编辑

The screenshot shows the 'Special Calculation' interface. On the left sidebar, under the 'Special Calculation' section, there is a red box around the '偏离度计算' (Deviation Calculation) component. A red arrow points from this box to a callout box labeled '1. 从左侧算法库中拖入偏离度计算组件' (Drag the deviation calculation component from the left algorithm library). Another red arrow points from the '偏离度计算' component in the main workspace to a callout box labeled '2. 在右侧参数栏中，可以通过点击添加比较，增加多组数据的计算' (In the right parameter bar, you can click '+ Add Comparison' to add more groups of data calculations). A third red arrow points from the '操作贴士' (Operation Tips) section to a callout box labeled '3. 点开可下拉选择需要计算的字段，以及设置目标数据字段' (Click to open and select the fields to be calculated and set the target data field). The main workspace displays a data flow diagram with components like '爱丽丝(E5iC1MB)', '字符转化', 'bob(cDG6bsY)', '缺失值填充', '联邦求交(FL)', and '偏离度计算'. The '操作贴士' section on the right contains tips about deviation formulas and ranges.

图135 偏离度计算使用说明

c) 运行结果

组件执行完成后，可右键查看计算结果，如下图所示。同时支持对计算结果的下载。

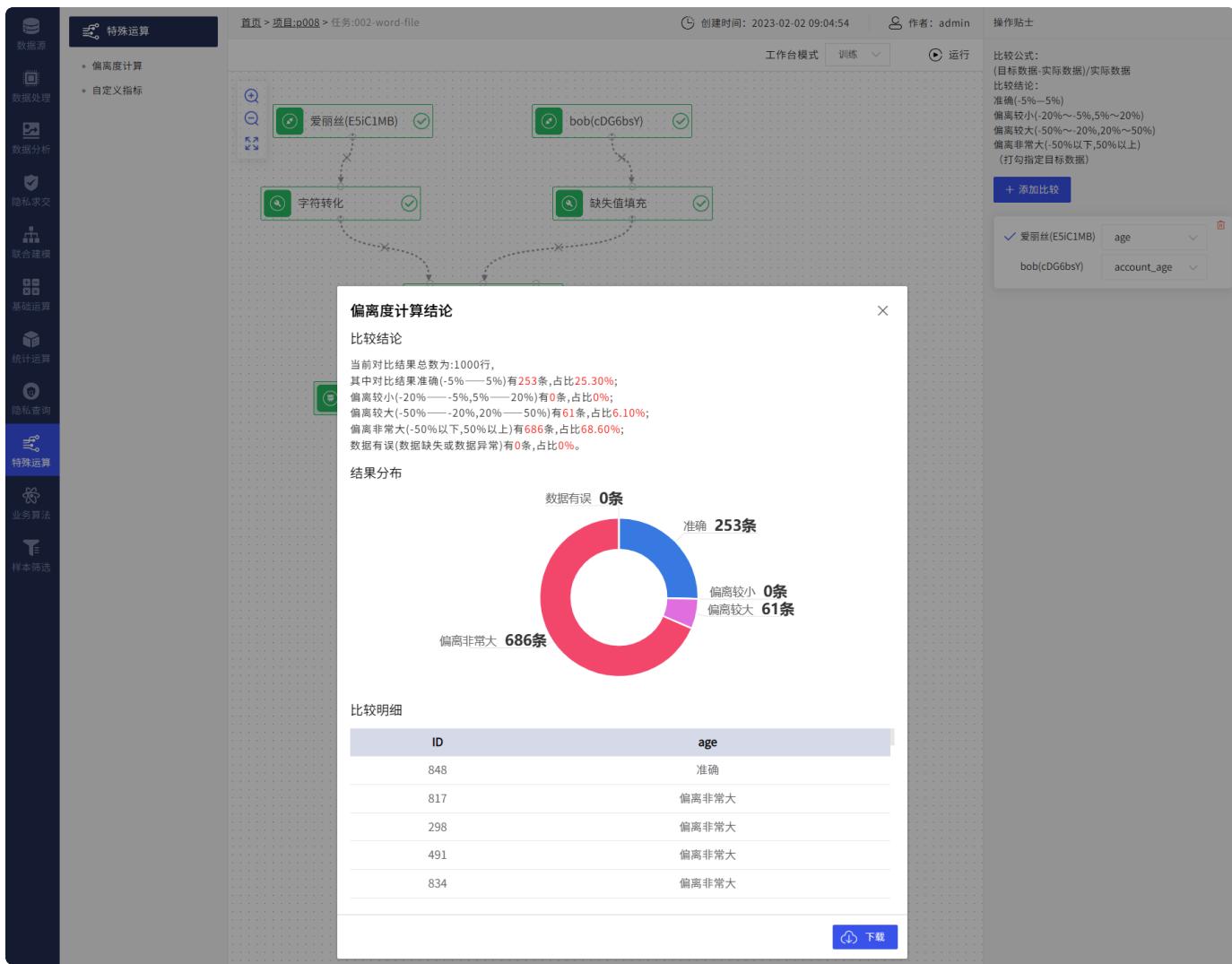


图136 偏离度计算结果

4.4.9.2. 自定义指标

a) 组件说明

基于多方安全技术实现的多方数据源之间的自定义四则运算（该组件需在求交后执行）。

b) 画布编辑

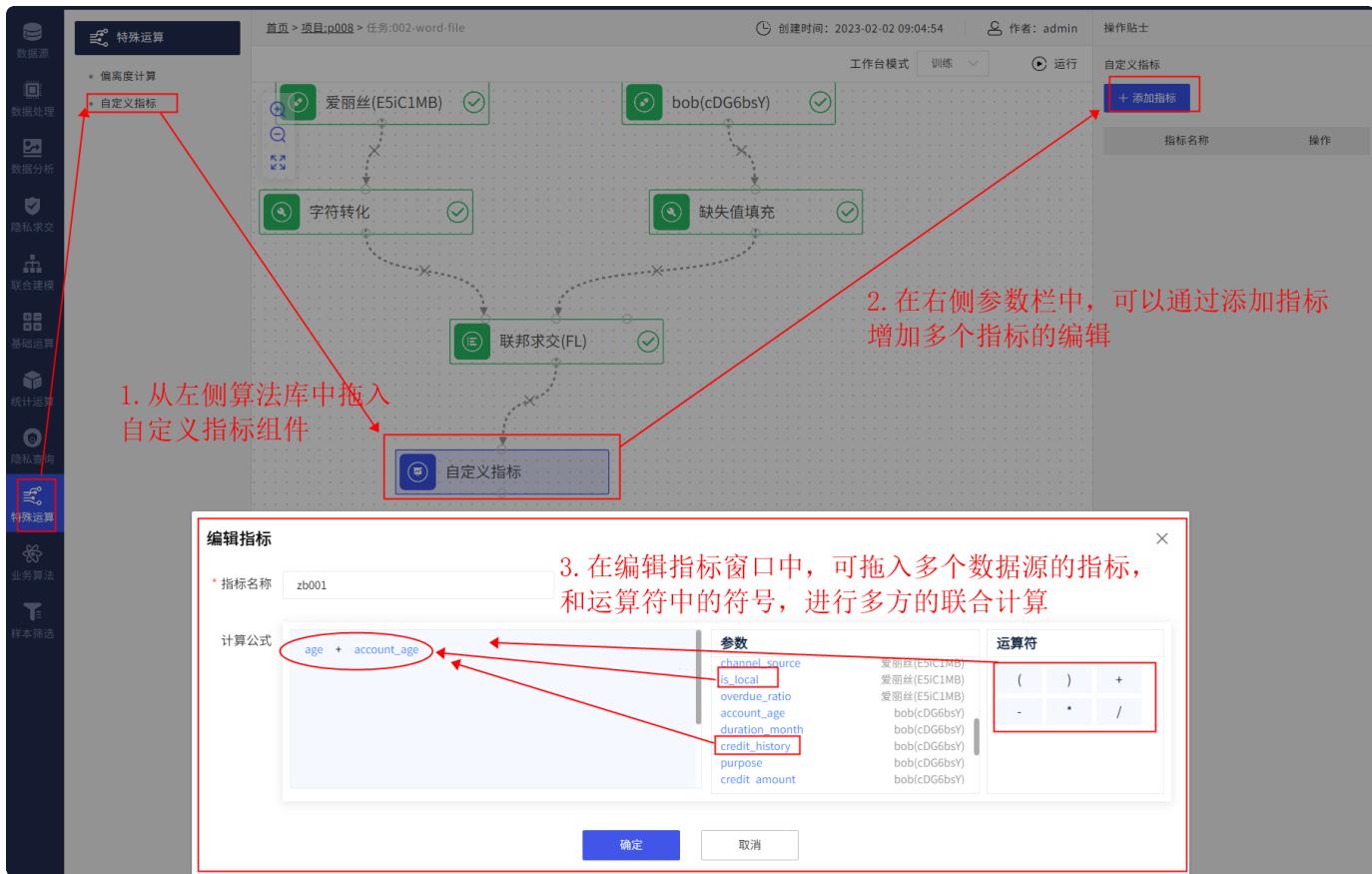


图137 自定义指标使用说明

c) 运行结果

组件执行完成后, 可右键查看计算结果, 如下图所示。同时支持对计算结果的下载。

The screenshot shows a data processing interface. On the left, a sidebar lists various operations: 特殊运算 (Special Operation), 偏离度计算 (Deviation Calculation), 自定义指标 (Custom Metrics), 数据源 (Data Source), 数据处理 (Data Processing), 数据分析 (Data Analysis), 隐私求交 (Privacy Intersection), 联合建模 (Joint Modeling), 基础运算 (Basic Operation), 统计运算 (Statistical Operation), 隐私查询 (Privacy Query), 特殊运算 (Special Operation) (selected), 业务算法 (Business Algorithm), and 样本筛选 (Sample Screening). The main area displays a workflow diagram with four nodes: 爱丽丝(E5iC1MB) (Alice), bob(cDG6bsY) (Bob), 字符转化 (Character Conversion), and 缺失值填充 (Missing Value Filling). Below the diagram is a modal window titled '自定义指标计算结果' (Custom Metric Calculation Result) with a red border. The table has columns 'ID' and 'zb001'. The data is as follows:

ID	zb001
848	2.0
817	5.0
298	5.0
491	3.0
834	4.0
225	5.0
238	5.0
359	2.0
830	6.0
127	3.0
189	3.0
175	5.0
470	3.0
994	5.0
292	3.0
774	4.0
674	6.0
135	5.0
353	2.0
705	5.0
841	5.0
529	3.0

计算结果 **自定义指标计算结果** **zb001**

图138 自定义指标计算结果

4.5. 项目编辑-模型应用

用户在完成联合建模任务后，可在项目列表中，点击该项目的模型应用，进入应用部署页面。

p008

在项目列表中，点击指定项目的模型应用

test_001

p001

p007

图 139 模型应用入口

通过选择指定的任务和指定的模型，以及对各个节点的应用集进行设置，可对指定的模型训练任务进行部署。

002 > 模型应用

部署模型

所属项目: 002

* 任务选择: 001-psi

* 模型选择: 逻辑回归

应用集设置: 金智塔科技(y5Qsn9t) > 已设置
bob(sh87fsK) > 已设置

1. 点击部署模型

2. 选择需要部署的任务和模型

3. 设置应用数据集

图140 部署模型

点击确定后，会生成一条部署模型信息。该部署模型信息可用于后续的模型应用。

The screenshot shows a deployment record for model ID 73337b2e05214820a8cfa991f231e167. The deployment time is 2023-02-20 11:11:16. The model type is Federated Learning (FL), and the name is Logical Regression. Model parameters include tol: 0.001, batch_size: 200, and C: 1.0. There are buttons for CSV application, database application, API application, execution process, and application log.

图141 模型部署记录列表

4.5.1. CSV应用

在完成部署后，用户可通过CSV应用进行查询ID数据集的上传。

The screenshot shows the CSV application interface. It has fields for sample ID format (CSV file selected), CSV file selection (choose file), and ID column name (please enter ID column name). A red box highlights the 'CSV Application' button, and a red arrow points to a callout box labeled '点击CSV应用可进行模型应用' (Click CSV application to perform model application).

图142 模型应用中的查询ID数据集设置

上传成功后，点击执行推断，即可进行模型预测的计算。计算完成后，可进行计算结果的查看，如下图所示。

001-psi
创建者 admin 更新时间 2023-02-20 10:17:12 需求节点 bob

应用集Feature设置：
[金智塔科技(ySQsn9t) > 已设置] [bob(sh87fsK) > 已设置]

模型ID	73337b2e05214820a8cfa991f231e167	模型类型	联邦学习 (FL)
部署时间	2023-02-20 11:11:16	模型名称	逻辑回归
		模型参数	tol: 0.001 batch_size: 200 C: 1.0 查看更多 >

CSV应用 数据库应用 API应用 执行过程 应用记录

共 1 条 5条/页 < 1 > 前往 1 页

CSV应用
样本ID形式
* 样本来源 CSV文件 数据库
* CSV文件 apply_id.csv
文件地址：/home/code/data/2919374418254115_824.csv
ID列名
执行推断 取消

1. 上传CSV文件成功
2. 设置文件中的ID字段名
3. 点击执行推断，即可进行模型应用

图143 查询ID数据集上传成功

001-psi
创建者 admin 更新时间 2023-02-20 10:17:12 需求节点 bob

应用集Feature设置：
[金智塔科技(ySQsn9t) > 已设置] [bob(sh87fsK) > 已设置]

模型ID	73337b2e05214820a8cfa991f231e167	模型类型	联邦学习 (FL)
部署时间	2023-02-20 11:11:16	模型名称	逻辑回归
		模型参数	tol: 0.001 batch_size: 200 C: 1.0 查看更多 >

CSV应用 数据库应用 API应用 执行过程 应用记录

共 1 条 5条/页 < 1 > 前往 1 页

应用记录

编号	执行时间	状态	应用集	应用方式	操作
98	2023-02-20 11:18:21	完成	V2_germancredit_part2 V2_germancredit_part1	CSV应用	查看结果

共 1 条 5条/页 < 1 > 前往 1 页

查看应用记录

图144 模型应用记录列表

点击查看结果，可在页面上查看应用结果。并可以对结果进行下载。

The screenshot shows the Jinzhita Technology Privacy Computing Platform interface. At the top, there is a navigation bar with the platform logo, user information (admin), and a '切换到管理台' (Switch to Management Console) button. Below the navigation bar, the page title is '模型应用' (Model Application). The main content area displays a model application record for '001-psi'. The record includes details such as '创建者 admin' (Creator: admin), '更新时间 2023-02-20 10:17:12' (Last updated: 2023-02-20 10:17:12), and '需求节点 bob'. It also shows the '模型ID' (Model ID: 73337b2e05214820a8cfa991f231e167), '模型类型' (Model Type: 联邦学习 (FL)), and '部署时间' (Deployment Time: 2023-02-20 11:11:16). The '模型参数' (Model Parameters) section lists 'tol: 0.001', 'batch_size: 200', and 'C: 1.0'. A 'Feature设置' (Feature Settings) section shows '金智塔科技(ySQsn9t)' and 'bob(jsh87fsK)' both set to '已设置' (Set). Below these details are five buttons: 'CSV应用' (CSV Application), '数据库应用' (Database Application), 'API应用' (API Application), '执行过程' (Execution Process), and '应用记录' (Application Record). At the bottom, there is a pagination bar showing '共 1 条' (1 item total), '5条/页' (5 items per page), and a single-page indicator.

点击查看结果，可查看模型应用结果，并支持下载

应用记录

编号98的应用结果
输出形式：CSV文件
文件名：models_f1_lr.csv
文件路径：result/0dxEQsQ2ZPMn7TL9tc8/0dxEQsQ2ZPMn7TL9tc8-models_f1_lrLPrQz4a/dataProvider-bob/models_f1_lr.csv

id	label	probability
1	1	83%
10	1	57%
2	0	17%
3	1	65%
4	1	79%

共 10 条 | 5条/页 | < 1 2 > | 前往 1 页

图145 模型应用结果查看

4.5.2. API应用

用户可通过API应用，获取模型预测结果。该模型结果为部署时，各参与方设置应用数据集后的模型预测结果。

The screenshot shows the Jinzhita Technology Privacy Computing Platform interface. At the top, there is a navigation bar with the platform logo, user information (admin), and a '切换到管理台' (Switch to Management Console) button. Below the navigation bar, a breadcrumb trail indicates the current location: '您现在所在的位置是: 002 > 模型应用'. On the right side of the header, there is a '部署模型' (Deploy Model) button.

The main content area displays a model configuration page for '001-psi'. Key details include:

- 创建者:** admin
- 更新时间:** 2023-02-20 10:17:12
- 需求节点:** bob
- 模型ID:** 73337b2e05214820a8cfa991f231e167
- 模型类型:** 联邦学习 (FL)
- 部署时间:** 2023-02-20 11:11:16
- 模型名称:** 逻辑回归
- 模型参数:** tol: 0.001, batch_size: 200, C: 1.0

Below the main configuration, there are several tabs: CSV应用, 数据库应用, API应用 (which is highlighted with a red box), 执行过程, and 应用记录. A red arrow points from the 'API应用' tab to the '请求地址 (api已就绪)' field in the API documentation panel.

The API application documentation panel contains the following sections:

- 请求地址 (api已就绪):** /platform/workflow/predict/694b0e26b9624922bf77cf53a4f74267 (highlighted with a red box)
- 请求示例:**
 - 请求示例1:**

```
{
  "applyType": 1 //1—id应用,2—id_list,
  "applyObj": 1 //待预测样本的id
}
```
 - 请求示例2:**

```
{
  "applyType": 2 //1—id应用,2—id_list,
  "applyObj": 1,2,3 //待预测样本id_list—id之间通过逗号连接
}
```

At the bottom of the API documentation panel, there are two buttons: '下载文档' (Download Document) and '关闭' (Close).

图146 API应用

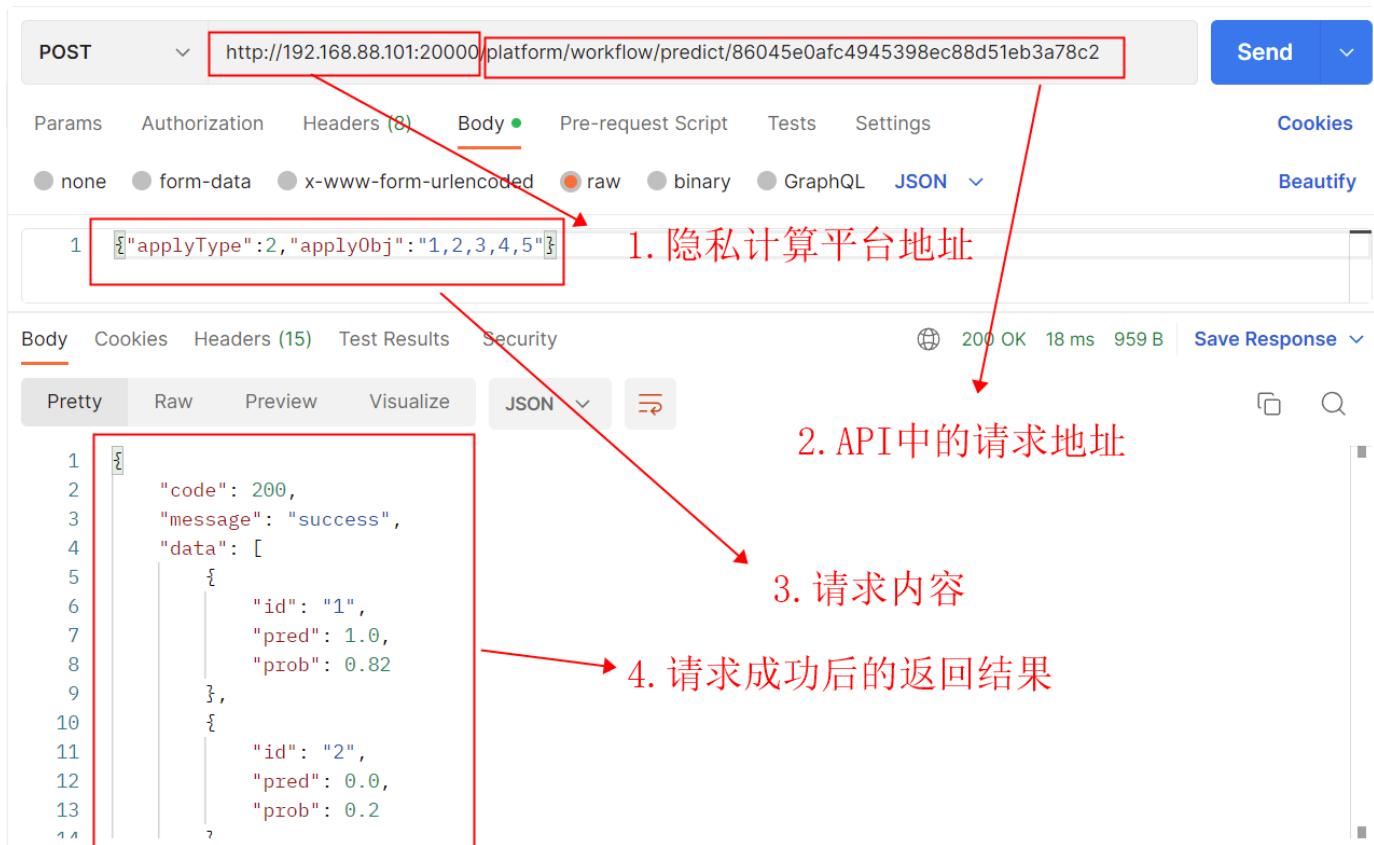


图147 postman请求

5.附录

5.1. 组件与参与方数量的兼容性

特别说明，同一个组件不支持同一个数据源的多个输入。如下图所示，追踪查询中有2个开发环境103、1个开发环境102，当前平台不支持此逻辑。

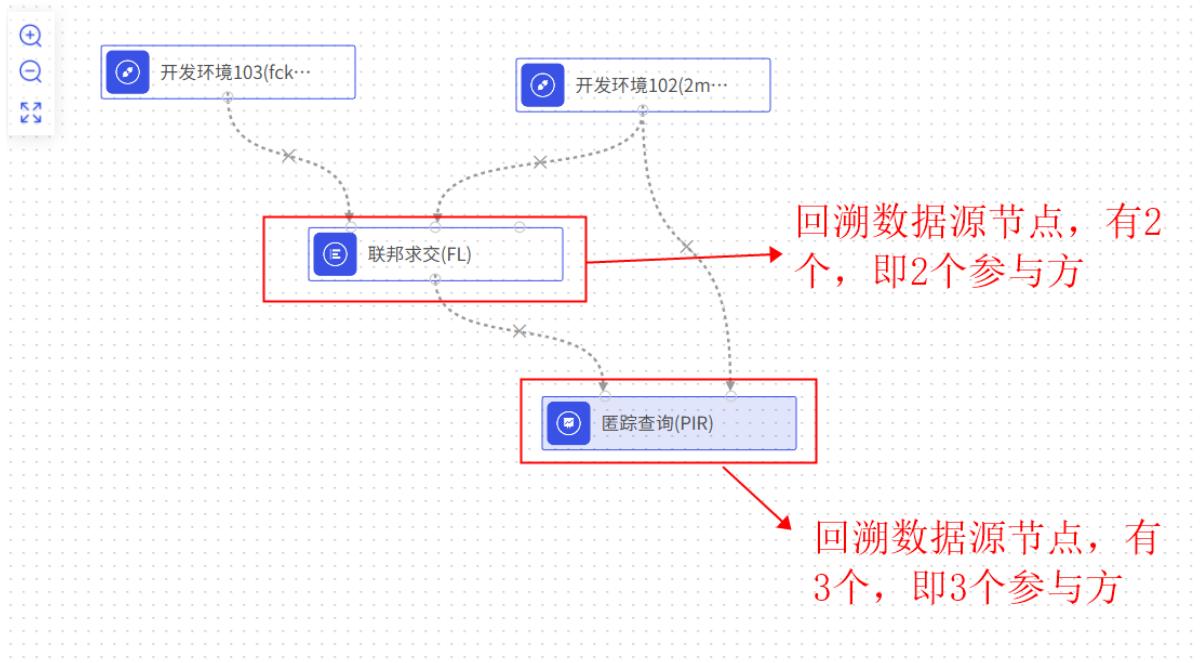


图148 参与方数量说明

表2 组件规则

组件类型	组件名称	支持参与方数量
数据处理	字符转化	>=1
	稀疏样本删除	>=1
	缺失特征删除	>=1
	缺失值填充	>=1
	特征分箱	>=1
	卡方分箱	>=2
	类编编码	>=1
	标准化	>=1
	Sql编辑	1
数据分析	Python编辑	>=1
	皮尔逊相关系数	2
	信息值IV	2,3
	特征重要性分析	>=2

	方差膨胀系数VIF	2
隐私求交	联邦求交	≥ 2
	多方安全求交	≥ 2
联合建模	逻辑回归	≥ 2
	Xgboost	≥ 2
	K-means	≥ 2
	逻辑回归 (MPC)	2,3
基础运算	安全加法	≥ 2
	安全减法	2
	安全乘法	2,3
	安全除法	2
	安全比较	2
统计运算	最大值	2,3
	中位数	2,3
	均值	≥ 2
	方差	2,3
隐私查询	匿踪查询	2
	加密传输	2
特殊运算	偏离度计算	2
	自定义指标	2,3