Smartbi V10.1-数据挖掘

新特性列表中：+表示新增；^表示增强；<表示变更

新增【+】	增强【^】
+【数据挖掘/ETL】关系数据源新增对华为Fusioninsight数据库的支持	^【ETL自动化】优化导入模板：实现模板的可视化配置
+【数据挖掘】新增一键推荐功能，用于推荐设置Spark 资源配置项的值	^【数据挖掘】关系数据源新增选择列功能，用于输出指定数据
+【数据挖掘】新增模型对比功能，支持算法模型可视化分析并导出评估报告	^【数据挖掘】关系目标表新建表支持添加字段别名到数据库
+【数据挖掘】作业流支持一次性多选节点拖拽到画布	^【数据挖掘/ETL】部分数据预处理方法功能增强
	^【数据挖掘】优化算法节点自动调参设置
	^【数据挖掘】优化全表统计节点
	^【数据挖掘】部署服务支持灰度部署方式
	^【数据挖掘】添加用户权限控制

^【ETL自动化】优化导入模板：实现模板的可视化配置

功能简介

新版本对ETL自动化主要是对模板进行了如下两点优化：可视化界面配置模板；由“TMPL_Data_Access_Configuration”和“TMPL_Data_Dictionary”两个模板简化成一个模板“ExcelTemplate”。

注意事项

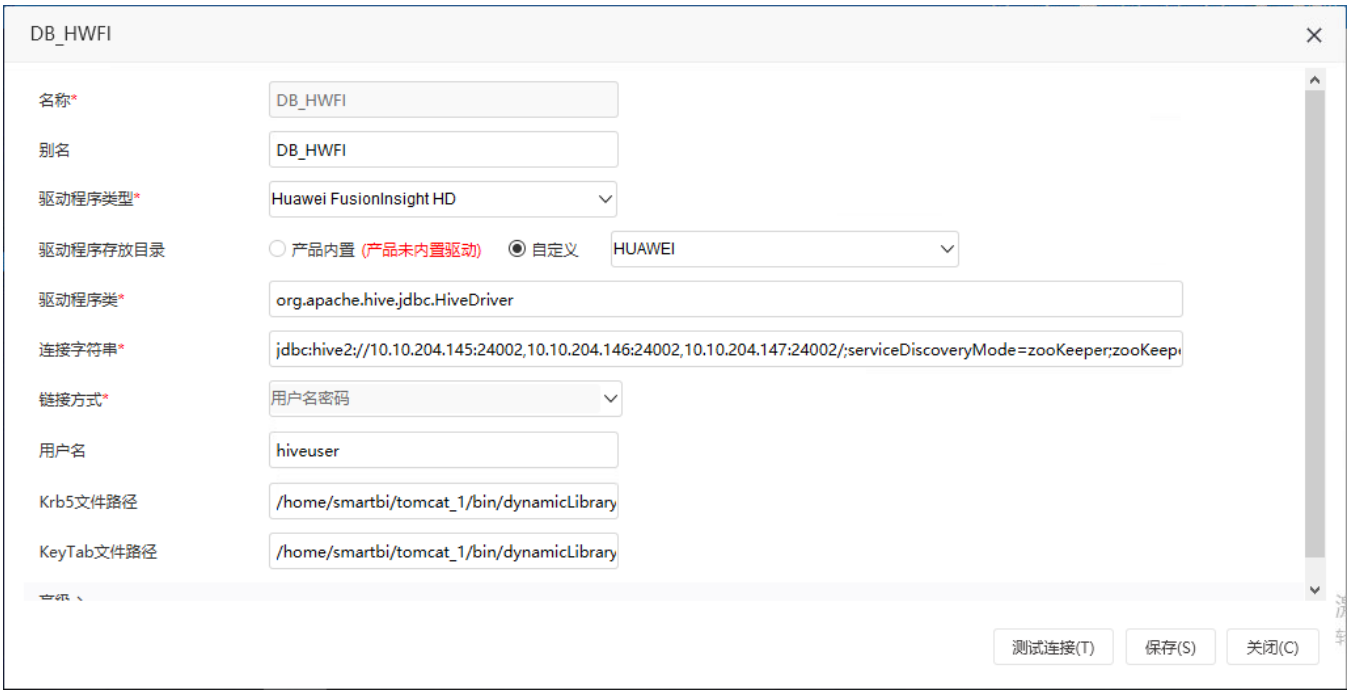
- 源数据库支持的类型有：Infobright、ClickHouse、Vertica、Oracle、MySQL、DB2、MSSQL、Presto、Hadoop_Hive、Guass100、PostgreSQL、Greenplum、星环、达梦、GBase、Sybase、HANA、Aliyun AnalyticDB（2.7.8版本）、ODPS。
- 目标数据库支持的类型有：ClickHouse、Oracle、Mysql、SqlServer。

+【数据挖掘/ETL】关系数据源新增对华为Fusioninsight数据库的支持

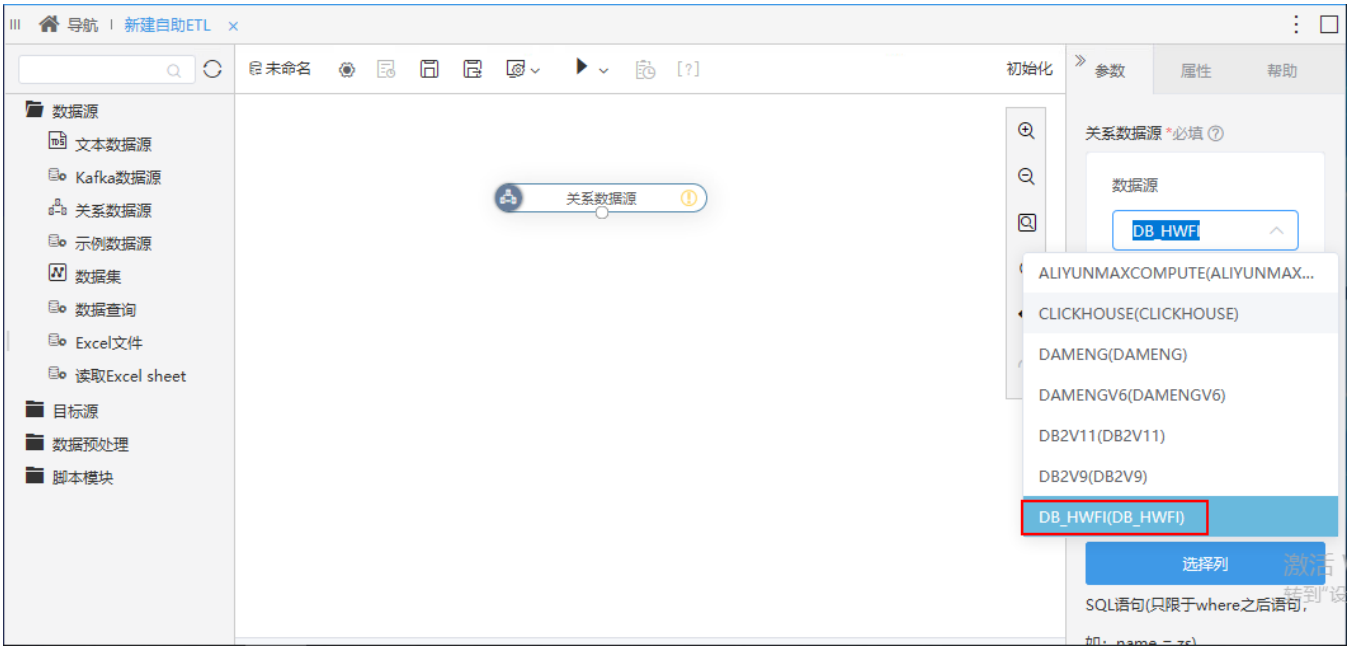
功能简介

新版本中，数据挖掘和ETL新增对华为Fusioninsight数据库的支持。

数据连接中配置好华为Fusioninsight数据库：



新建实验或ETL时，可以选择该数据源：



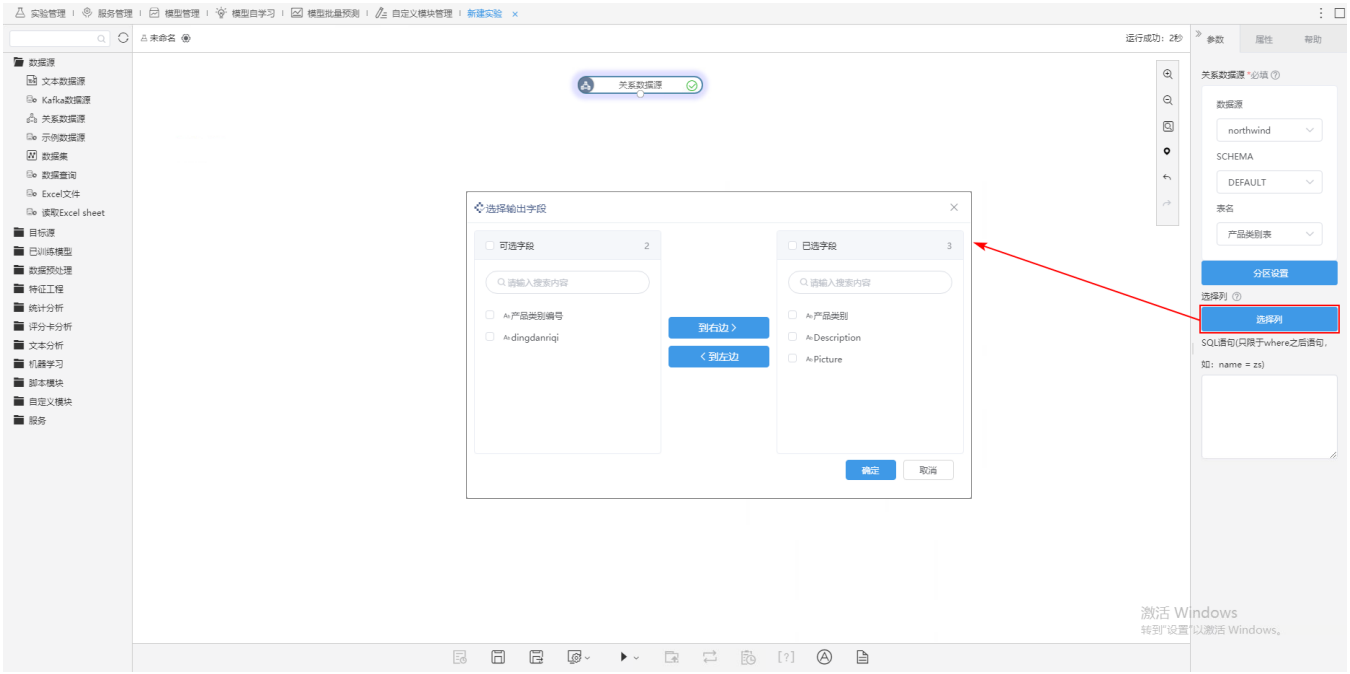
^【数据挖掘】关系数据源新增选择列功能，用于输出指定数据

背景介绍

在实际应用中，由于关系数据源导出数据时无法选字段，只能导入全部字段后通过列选择筛选字段，如果导入的数据过多，会导致运行时间过长。新版本，在数据模型的ETL高级查询、自助ETL、数据挖掘中，关系数据源新增选择列功能，可用于输出指定的数据，提升数据导出效率。

功能简介

新版本，在数据模型的ETL高级查询、自助ETL、数据挖掘中，关系数据源新增选择列功能，可用于输出指定的数据。



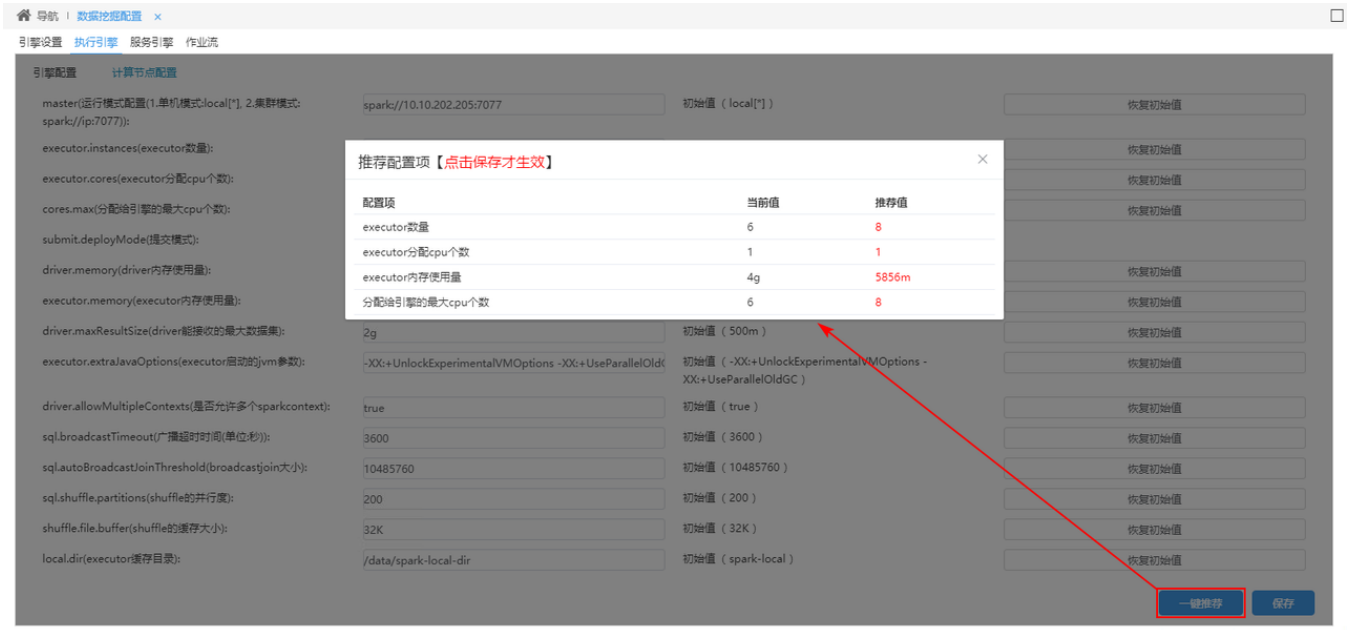
+【数据挖掘】新增一键推荐功能，用于推荐设置Spark 资源配置项的值

背景介绍

Spark是一种快速、通用、可扩展的大数据分析引擎 ， 能够高效地支持更多计算模式 ， 适用于各种各样原先需要多种不同的分布式平台的场景等。为了帮助用户更好的配置Spark 资源，新版本在数据挖掘配置中新增“一键推荐”功能，用于推荐设置Spark资源配置项的值。

功能简介

新版本，在 **数据挖掘配置>执行引擎>计算节点配置** 页面，新增“一键推荐”功能，系统会根据Spark work节点的服务器资源，生成推荐的配置。



参考文档

详情请参考 Smartbi连接Spark 。

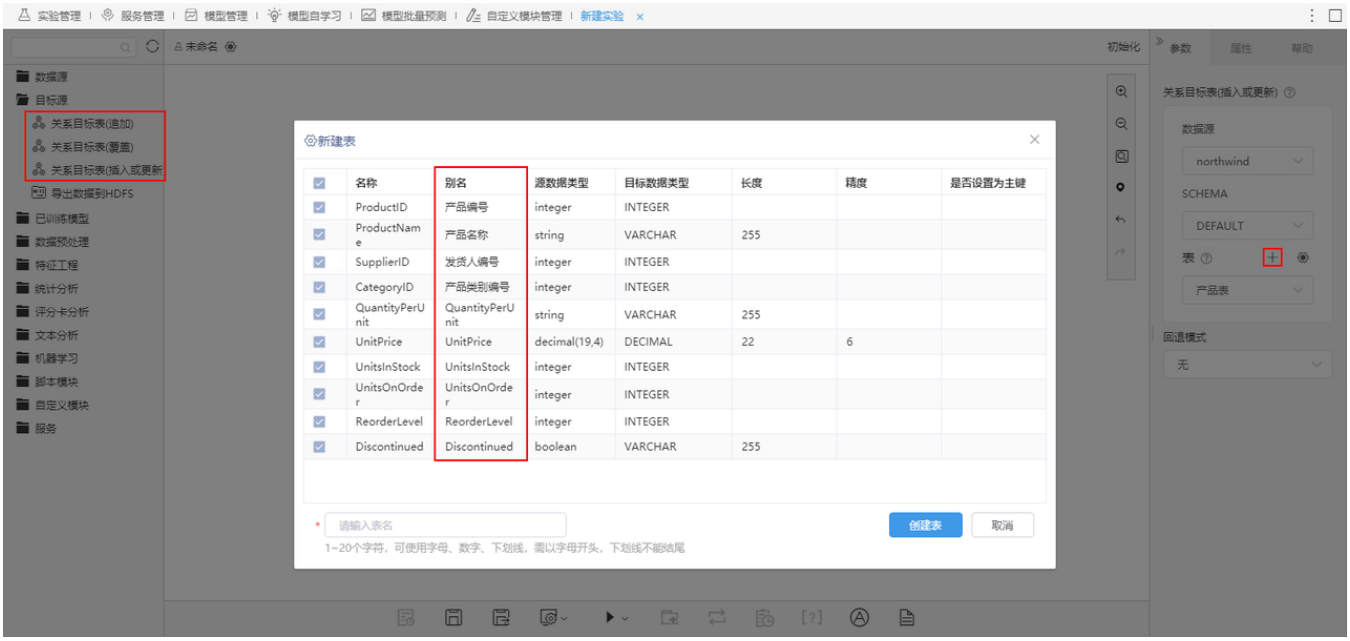
^【数据挖掘】关系目标表新建表支持添加字段别名到数据库

背景介绍

在实际应用中，有的用户在数据源表或使用元数据编辑节点修改了字段别名，由于数据库中没有别名属性，导致在实验中无法添加修改的别名。为了解决以上问题，新版本在关系目标表三个节点中新建表时，支持添加字段别名到数据库。

功能简介

新版本，在关系目标表三个节点中新建表时，支持添加字段别名到数据库。



注意事项

不支持此功能的数据源有：DAMENG_V6，MSSQL，VERTICA，XINGHUAN，GBASE8T，GBASE8S_V84。详情请参考 关系目标表 。

^【数据挖掘/ETL】部分数据预处理方法功能增强

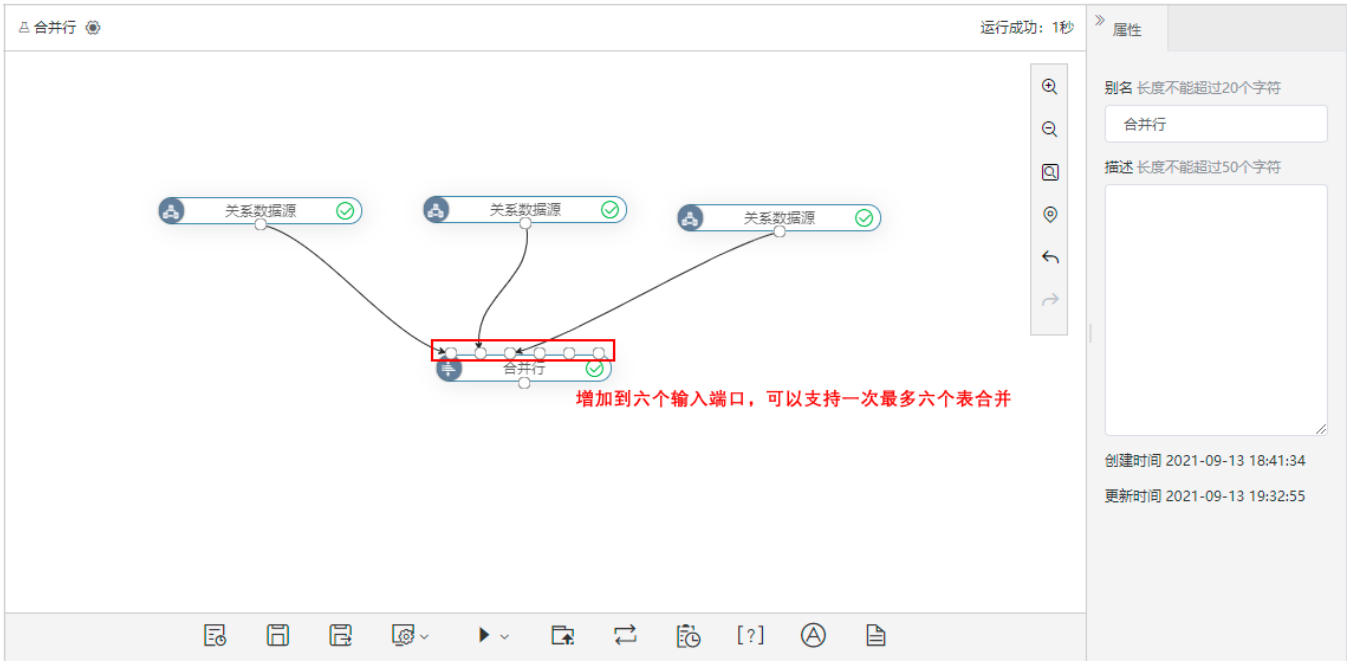
背景介绍

在ETL项目实施过程中，提出了希望能支持对多列字段设置不同的空值处理规则，以及多表合并的需求。因此，新版本基于项目需求，增强了“空值处理“及“合并行”的预处理方法。

功能简介

空值处理策略增强如下：

合并行增强如下：



参考文档

空值处理策略详情请参见：数据挖掘-空值处理。

合并行详情请参见：数据挖掘-合并行。

+【数据挖掘】新增模型对比功能，支持算法模型可视化分析并导出评估报告

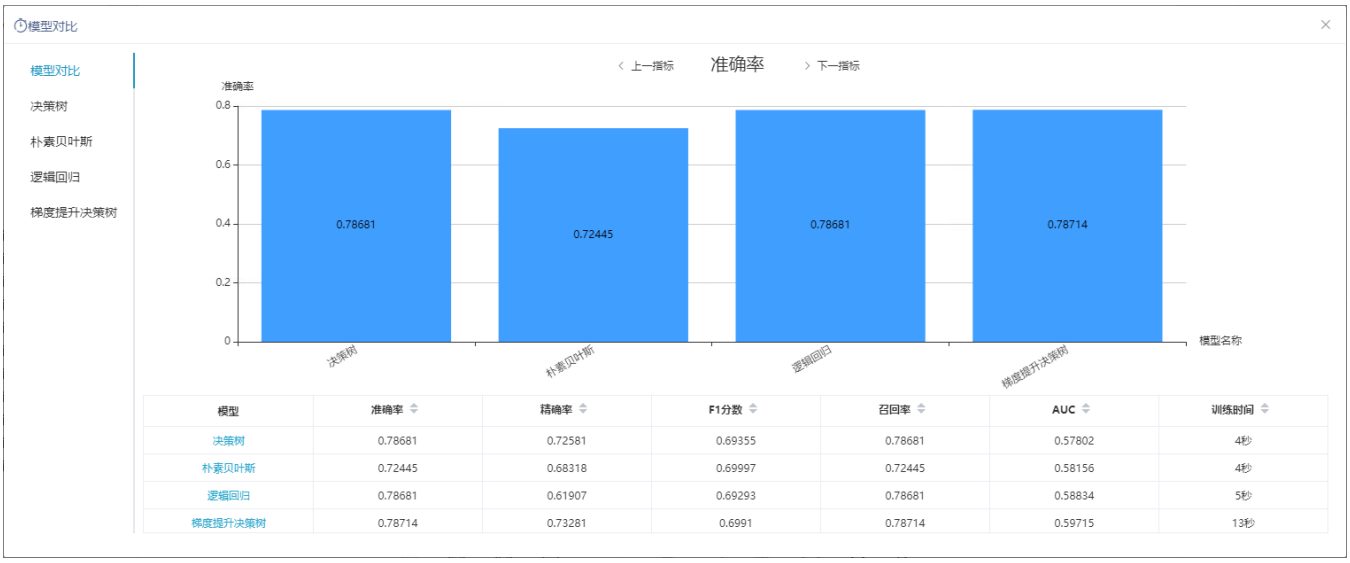
背景介绍

新版本，数据挖掘新增模型对比功能，支持对算法模型进行可视化分析并导出评估报告，可带来以下好处：

- 支持以可视化的方式对模型训练进行评估和对比，方便业务人员更快确定最佳预测模型，提供高可信度的答案；
- 对于效果不同的模型，支持将分散的评估节点的各种参数和结果进行汇总，便于业务人员查看和保存。

功能简介

新版本，数据挖掘支持模型对比功能，支持对算法模型进行可视化分析并导出评估报告。



参考文档

详情请参考 数据挖掘-评估报告 。

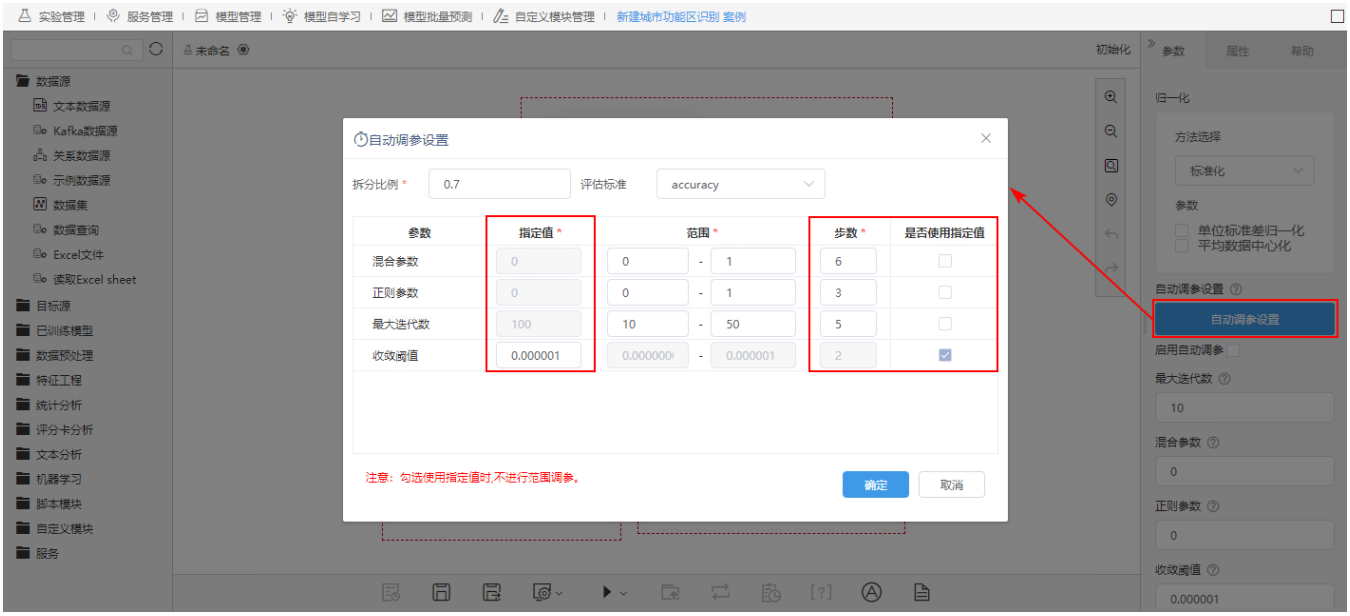
^【数据挖掘】优化算法节点自动调参设置

背景介绍

在实际应用中，设置算法的调参的范围时，大多数情况下并不需要对所有的参数进行调参，并且对于不同的参数，其需要的可选值的个数也不一样。新版本，算法节点的自动调参设置中新增指定值、步数等设置项，可减少自动调参的时间。

功能简介

新版本，支持向量机、梯度提升决策树、逻辑回归、线性回归节点的自动调参设置中，新增”指定值“、”步数”、”是否使用指定值”设置项。



其中，支持向量机的自动调参设置中，新增了“AUC（二分类）”评估标准。

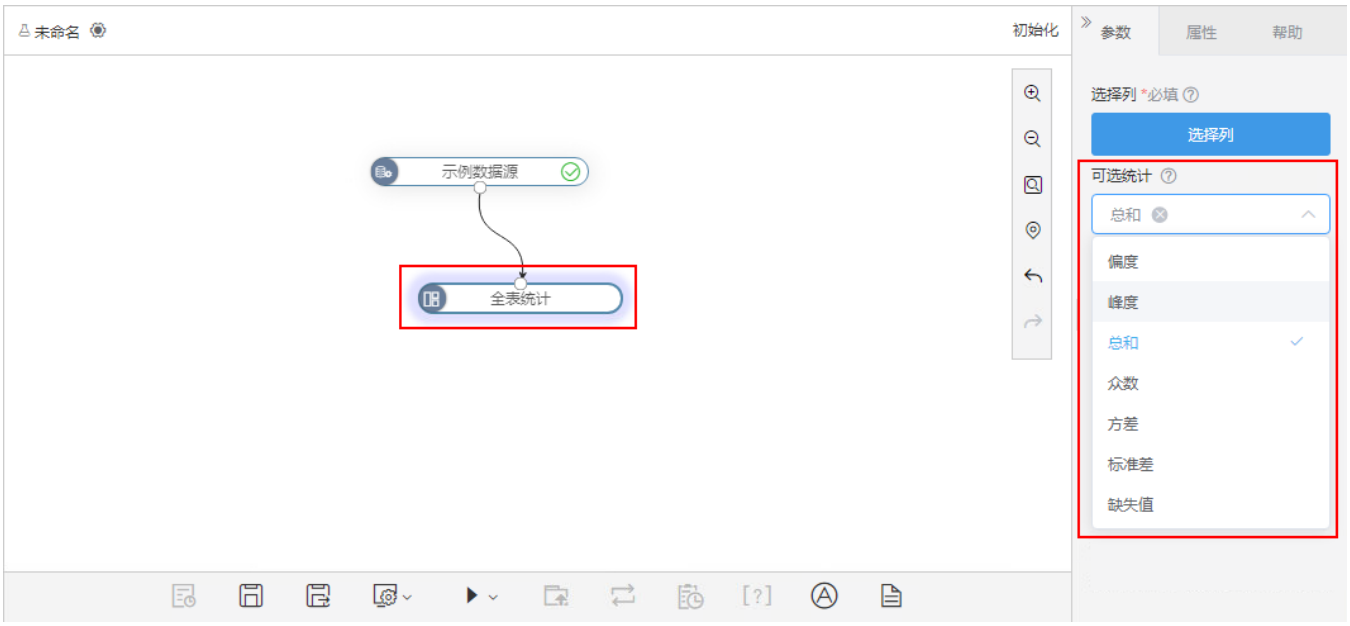
^【数据挖掘】优化全表统计节点

背景介绍

在实际应用中，用户使用统计分析时，如果观测数据包含的字段较多，各项指标的统计时间会很长，用户需要花费很多时间等待运行结果。新版本，Smartbi将全表统计的指标拆分为默认指标和可选的指标，用户在运行实验时无需运行全部指标，可极大地减少运行时间，提升性能。

功能简介

新版本，全表统计新增“可选统计”设置项，用户可根据需求添加对应的指标进行统计分析。



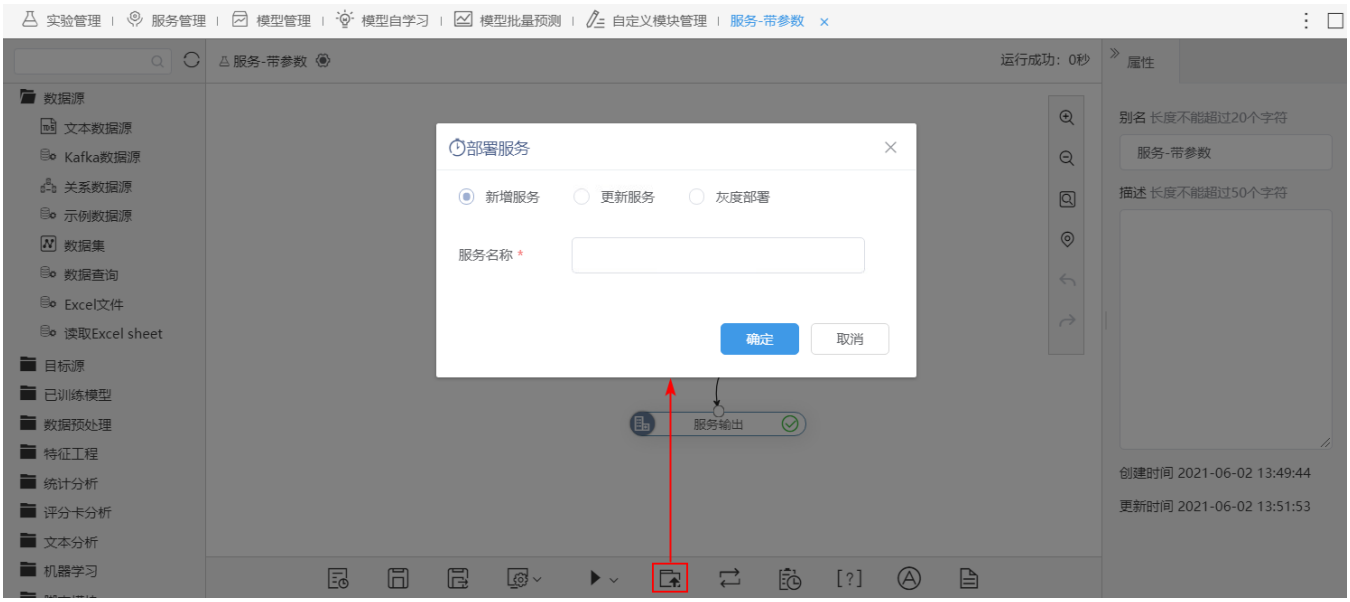
^【数据挖掘】部署服务支持灰度部署方式

背景介绍

新版本，数据挖掘部署服务中新增灰度部署功能，用于使多版本模型并行运行，并保持多版本模型运行的结果，通过结果比较灰度测试的模型版本的准确性和稳定性。

功能简介

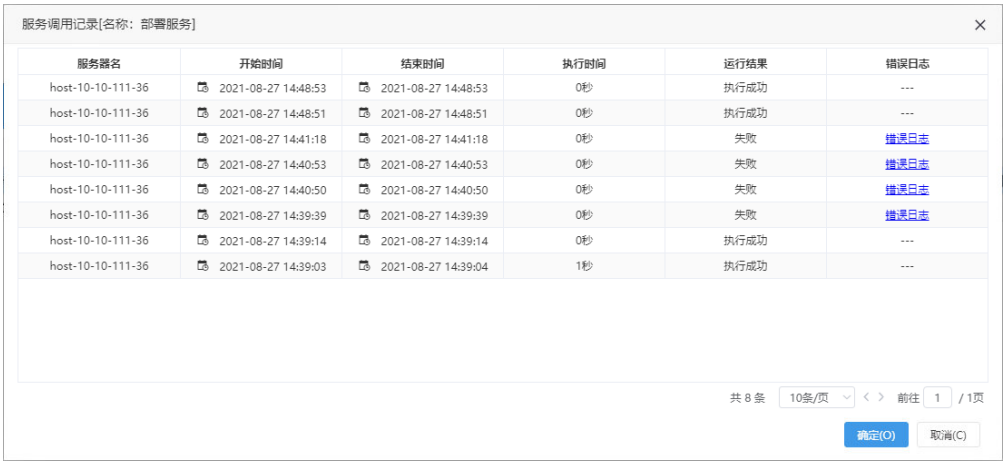
1、新增灰度部署功能，并修改部署方式为：新增服务、更新服务、灰度部署。



2、服务管理中新增服务调用统计、服务调用记录：

- 服务调用统计用于对某个时间段服务调用执行记录进行统计。

- 服务调用记录记录了服务调用的服务器名、开始时间、结束时间、错误日志等信息，可用于监控模型的运行时长、稳定性等。



参考文档

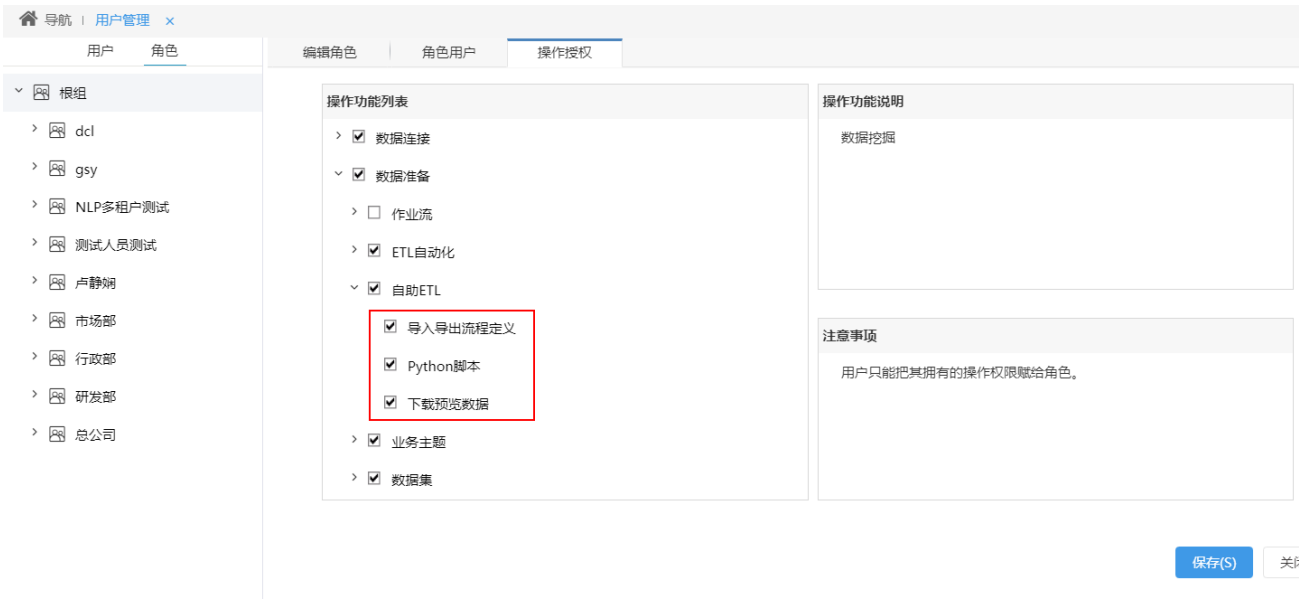
详情请参考 数据挖掘-服务 。

^【数据挖掘】添加用户权限控制

功能简介

新版本，数据模型ETL高级查询、自助ETL、数据挖掘中添加用户权限控制，限制用户能否查看评估报告、导入和导出流程图、使用Python脚本节点、下载预览数据功能。

- 自助ETL和数据模型ETL高级查询通过以下设置项控制对应权限：



- 数据挖掘通过以下设置项控制对应权限：

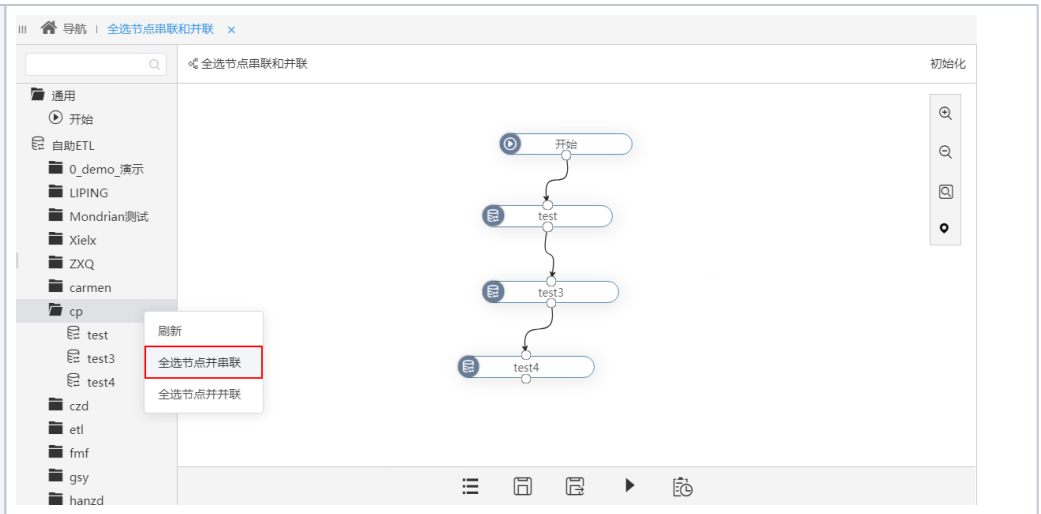
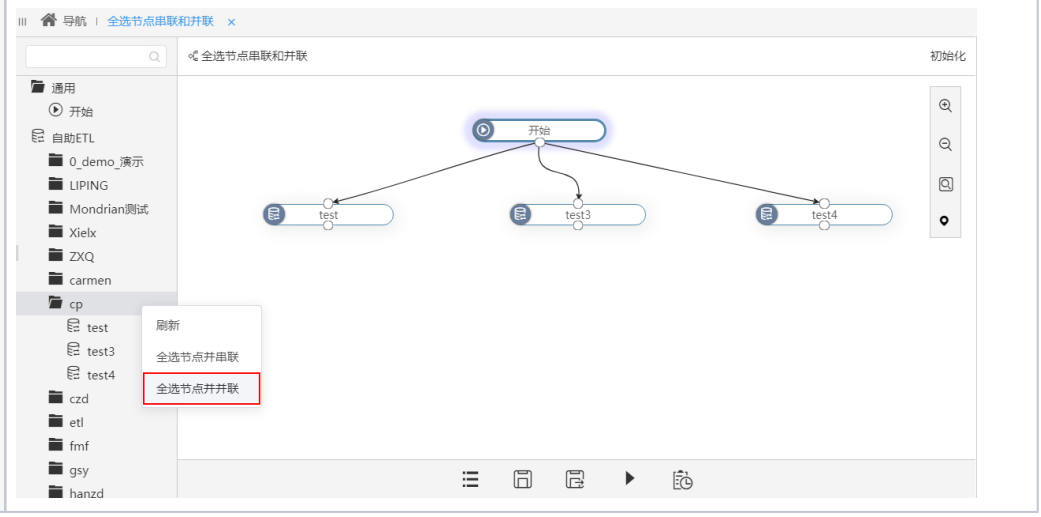
+【数据挖掘】作业流支持一次性多选节点拖拽到画布

背景介绍

在实际应用中，用户有时候会一次生成几百个自助ETL，因此在新建作业流时节点的数量会很多，如果这个时候手动一个个去拖拽，工作量庞大且容易出错。新版本，在作业流中支持“全选节点串联”和“全选节点并联”功能，可快速将选择的目录中所有的节点以串联或并联的方式在画布区中连接，减少了重复操作，提高工作效率。

功能简介

新版本，作业流中支持“全选节点串联”和“全选节点并联”功能，可快速将选择的目录中所有的节点以串联或并联的方式在画布区中连接。

串联	 导航 全选节点串联和并联 初始化 通用 开始 自助ETL 0_demo_演示 LIPING Mondrian测试 Xielx ZXQ carmen cp test test3 test4 czd etl fmf gsy hanzd 刷新 全选节点并串联 全选节点并并联 开始 test test3 test4
并联	 导航 全选节点串联和并联 初始化 通用 开始 自助ETL 0_demo_演示 LIPING Mondrian测试 Xielx ZXQ carmen cp test test3 test4 czd etl fmf gsy hanzd 刷新 全选节点并串联 全选节点并并联 开始 test test3 test4