

DTEmpower V2022R2

用户操作手册

2022.07.29



目 录

1	软件简介	1
1.1	定位及功能	1
1.2	功能特色	2
1.2.1	丰富而先进的算法工具箱	2
1.2.2	零编码建模	2
1.2.3	基于表格视图与交互的数据前处理	3
1.2.4	智能数据清理算法	3
1.2.5	针对小规模数据集的 AI Agent 训练	4
1.2.6	机理模型融合	5
1.2.7	工业设计场景聚焦——智能优化	5
1.2.8	工业运维场景聚焦——智能监盘	6
2	项目组成	7
2.1	层级结构划分	7
2.2	项目	7
2.3	工程	8
2.4	数据管理	8
2.5	数据源管理	10
2.6	建模流程	11
2.6.1	节点	12
2.6.2	节点连接	16
2.7	模型管理	16
2.7.1	模型响应面信息	17
2.7.2	模型应用	17
3	界面布局	18
3.1	Ribbon 菜单	19
3.1.1	文件	19
3.1.2	数据处理	19
3.1.3	建模流程	21
3.1.4	模型处理	22
3.2	资源树	23
3.3	建模区	24



3.3.1	表格式建模区.....	24
3.3.2	拖拽式建模区.....	25
3.4	底边栏.....	25
4	基础操作说明	26
4.1	软件安装.....	26
4.1.1	安装启动程序.....	26
4.1.2	浏览安装路径.....	27
4.1.3	文件解压与拷贝.....	27
4.1.4	以管理员身份运行程序.....	28
4.1.5	许可请求.....	29
4.1.6	许可配置.....	30
4.2	软件启动.....	30
4.2.1	客户端启动.....	31
4.2.2	本地浏览器接入.....	31
4.3	软件欢迎页.....	32
4.3.1	新建项目.....	33
4.3.2	打开项目.....	33
4.3.3	最近项目.....	34
4.3.4	示例项目.....	34
4.3.5	使用帮助.....	35
4.3.6	问题反馈.....	35
4.4	项目管理.....	35
4.4.1	新建项目.....	35
4.4.2	保存项目.....	36
4.4.3	打开项目.....	37
4.5	数据管理.....	37
4.5.1	上传数据集.....	37
4.5.2	新建数据集.....	39
4.5.3	复制数据集.....	39
4.5.4	下载数据集.....	40
4.5.5	重命名数据集.....	40
4.5.6	删除数据集.....	41
4.5.7	清空数据管理.....	41



4.6	数据前处理	42
4.6.1	加载全部	42
4.6.2	通用表格操作	43
4.6.3	更新与撤回	47
4.6.4	变量格式	47
4.6.5	变量核查	53
4.6.6	数据赋值	56
4.6.7	数据合并	61
4.7	数据可视化	64
4.7.1	1D 箱线图绘制	64
4.7.2	2D 散点图绘制	65
4.7.3	3D 散点图绘制	66
4.7.4	高维可视化绘制	66
4.7.5	相关图绘制	68
4.7.6	平行图绘制	68
4.7.7	柱状图绘制	69
4.7.8	折线图绘制	69
4.8	数据可视化交互配置	69
4.8.1	标题设置	70
4.8.2	x 轴变量设置	70
4.8.3	y/z 轴变量设置	71
4.8.4	新增副 y/z 轴	73
4.8.5	删除副 y/z 轴	73
4.8.6	共用 y 轴	74
4.8.7	图例设置	76
4.8.8	区域缩放与还原	76
4.8.9	数据聚类可视化	77
4.9	数据源管理	78
4.9.1	连接数据源	78
4.9.2	新建数据源	81
4.9.3	保存数据源	81
4.9.4	编辑数据源	81
4.9.5	删除数据源	81



4.10	建模流程	82
4.10.1	加入流程	82
4.10.2	新建分析	82
4.10.3	添加节点	83
4.10.4	节点属性配置	83
4.10.5	节点连线	84
4.10.6	删除连线	85
4.10.7	节点间输入输出量关联	85
4.10.8	多选节点	86
4.10.9	删除节点	86
4.10.10	自定义工具箱	87
4.10.11	执行流程	89
4.10.12	运行日志查看/关闭	89
4.10.13	新建流程	90
4.10.14	管理流程	91
4.10.15	终止流程	91
4.10.16	节点执行结果信息查看	92
4.11	模型管理	93
4.11.1	保存模型	93
4.11.2	下载模型	93
4.11.3	上传模型	94
4.11.4	模型信息查看	94
5	分析节点	96
5.1	数据载入	97
5.1.1	数据读取	97
5.1.2	数据采样	99
5.2	数据设定	100
5.2.1	变量设定	100
5.2.2	数据分割	102
5.3	数据筛选	103
5.3.1	变量范围	103
5.3.2	变量筛选	105
5.3.3	变量剔除	107



5.3.4	稀释采样.....	108
5.3.5	空值处理.....	110
5.3.6	自定义剔除.....	111
5.4	表格处理.....	112
5.4.1	重命名.....	112
5.4.2	数据存储.....	114
5.5	时序数据操作.....	115
5.5.1	时序变量设定.....	115
5.5.2	时序预处理.....	118
5.6	数据清理.....	122
5.6.1	智能清理算法 AIOD.....	122
5.6.2	基元算法清理.....	125
5.7	数据聚类.....	127
5.7.1	KMeans.....	127
5.7.2	Batch-KMeans.....	129
5.7.3	dr_gmm.....	130
5.7.4	fcm.....	132
5.7.5	Mean-Shift.....	133
5.7.6	normal_gmm.....	135
5.8	数据降维.....	137
5.8.1	PCA.....	137
5.8.2	batch_PCA.....	139
5.8.3	poly_KPCA.....	141
5.8.4	rbf_KPCA.....	142
5.8.5	sigmoid_KPCA.....	144
5.8.6	cosine_KPCA.....	146
5.8.7	FA.....	147
5.8.8	DL.....	149
5.8.9	LSA.....	151
5.8.10	ICA.....	152
5.8.11	isomap.....	154
5.8.12	LLE.....	156
5.8.13	LTSA.....	158



5.8.14	hessian_LLE.....	159
5.8.15	modified_LLE.....	161
5.9	数据变换.....	163
5.9.1	MinMax.....	163
5.9.2	MaxAbs.....	164
5.9.3	Z-score.....	166
5.9.4	Robust.....	168
5.9.5	Multiplication.....	170
5.9.6	Norm-1.....	172
5.9.7	Norm-2.....	173
5.10	自定义提取.....	175
5.10.1	表达式.....	175
5.10.2	数值运算-一元运算.....	177
5.10.3	数值运算-二元运算.....	179
5.10.4	数值运算-多元运算.....	181
5.11	相关性.....	183
5.11.1	GRA.....	183
5.11.2	Correlation.....	186
5.11.3	MIC.....	188
5.12	重要性.....	190
5.12.1	MDA.....	190
5.12.2	MDI.....	192
5.13	AI Agent.....	195
5.13.1	AI Agent.....	195
5.14	线性拟合.....	196
5.14.1	BayesianRidge.....	196
5.14.2	ElasticNet.....	198
5.14.3	ElasticNetCV.....	199
5.14.4	Huber.....	201
5.14.5	LAR.....	202
5.14.6	Lasso.....	204
5.14.7	LassoLar.....	205
5.14.8	ARD.....	207



5.14.9	Linear	208
5.14.10	Ridge	210
5.14.11	SGD	211
5.15	非线性拟合	213
5.15.1	AdaBoost	213
5.15.2	BaggingDeFo	215
5.15.3	CART	217
5.15.4	ExtraTrees	219
5.15.5	GP	221
5.15.6	GBDT	223
5.15.7	HistGBDT	225
5.15.8	KernelRidge	227
5.15.9	KNN	229
5.15.10	MLP	231
5.15.11	RNN	233
5.15.12	RandomForest	235
5.15.13	SVR	237
5.15.14	EnsembleRBF	239
5.15.15	RBFSurrogate	240
5.15.16	多项式拟合	241
5.16	调度策略	243
5.16.1	EnsembleBagging	243
5.16.2	HierachicalStratify	245
5.16.3	OTSpecifiedSDASRegressor	246
5.17	机理模型	248
5.17.1	显函数	248
5.18	时序训练	250
5.18.1	ARIMA	250
5.18.2	SARIMA	253
5.18.3	ES	256
5.18.4	LOESS	258
5.18.5	Prophet	260
5.18.6	TBATS	262



5.19	模型操作	265
5.19.1	模型聚合	265
5.19.2	模型对比	266
5.19.3	模型存储	267
5.19.4	模型读取	268
5.19.5	模型计算	269
5.19.6	批量训练	270
5.20	时序模型操作	272
5.20.1	模型预测	272
5.20.2	时序模型更新	273
5.20.3	时序模型读取	274
6	应用案例	275
6.1	简单案例 1——智能数据挖掘	275
6.1.1	案例描述	275
6.1.2	操作流程	277
6.1.3	效果对比	285
6.2	简单案例 2——专家知识耦合	287
6.2.1	案例描述	287
6.2.2	操作流程	288
6.2.3	效果对比	296
6.3	简单案例 3——数据前处理与可视化	297
6.3.1	案例描述	297
6.3.2	操作流程	298
6.4	专业案例 4——智能模型预警	304
6.4.1	案例描述	304
6.4.2	操作流程	305
7	注意事项	313
7.1	浏览器兼容性问题	313
7.2	表头要求	313
7.3	数据要求	313
7.4	表达式函数列表	314



1 软件简介

1.1 定位及功能

工业企业均拥有大量的数据，如何从数据中挖掘知识，并有效服务于自身的业务是各企业非常关注的话题。随着人工智能和数据挖掘技术的发展，市场上已存在有大量开源的和商业的数据建模解决方案，但是对于工业企业，要想建立高质量的模型，并应用于自己的业务场景，仍存在着一定的门槛，数据荒废或投入产出失衡的现象屡见不鲜。在此背景之下，南京天沭软件有限公司通过对工业企业数据建模需求的深度挖掘，研发了一套针对工业用户的数据建模软件 DTEmpower。

数据建模软件 DTEmpower 围绕数据清理、特征生成、敏感性分析和模型训练等数据建模的各个环节，提供有大量算法。通过针对特定场景下算法的深度研发，利用智能调度引擎和超参优化等技术，提高模型质量的同时，降低了对用户数据建模经验的需求。同时 DTEmpower 提供一套图形化的建模开发环境，所有算法均可通过拖拽的方式进行调用，通过连线的方式进行数据的传递，极大的降低了用户的使用门槛。依托强大的算法和简便的操作，零基础用户也可以利用 DTEmpower 快速挖掘到优秀的模型。

模型是可复用的知识，DTEmpower 秉承这一理念，定义了一套模型交换格式（简称 DT 模型），挖掘得到的模型可以直接存储为单个模型文件，DT 模型可以在 DTEmpower 的模型运行模块 DTEmpower Run（简称 DTRun）、天沭智能优化设计平台 AIPOD 等软件中直接调用，服务于实时预警和优化设计等各种应用。

“降低数据建模的门槛、强化模型的知识属性”，DTEmpower 针对典型的工业应用场景，提供了从建模到模型管理应用的一站式解决方案，让工业用户可以聚焦于业务而不是疲于数据分析，充分发挥数据的价值。



1.2 功能特色

1.2.1 丰富而先进的算法工具箱

DTEmpower 的核心功能建立于丰富、先进的算法工具箱基础之上。围绕数据清理、特征生成、敏感性分析和模型训练等数据建模的各个环节，DTEmpower 均提供有丰富的算法控件。在此基础上，“算法有自研，超参有优化，组合有调度，场景有聚焦”，相比于市面上的其他同类算法，DTEmpower 的模型训练效果精度更高、稳定性更好。

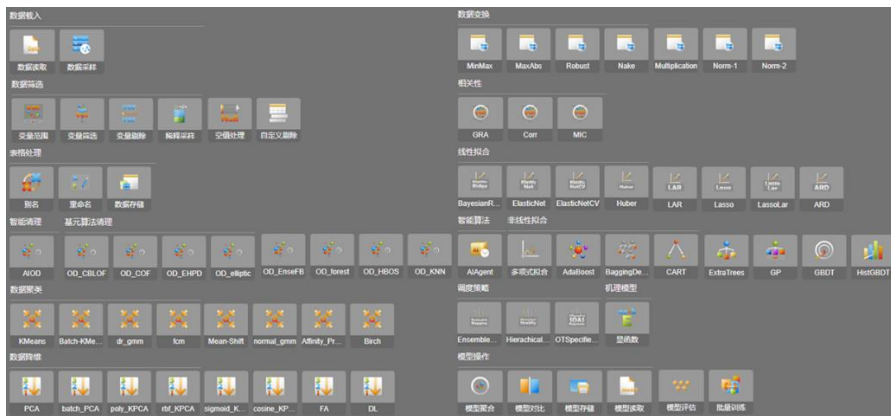


图1 DTEmpower 内置丰富而先进的算法工具箱

1.2.2 零编码建模

DTEmpower 提供了图形化的数据建模流程搭建功能。所有的数据及模型操作，均以工具箱中模块的形式提供。用户无需具备编码能力或深入了解算法细节，通过简单的节点拖拽与节点连接即可完成复杂的数据建模流程的构建。

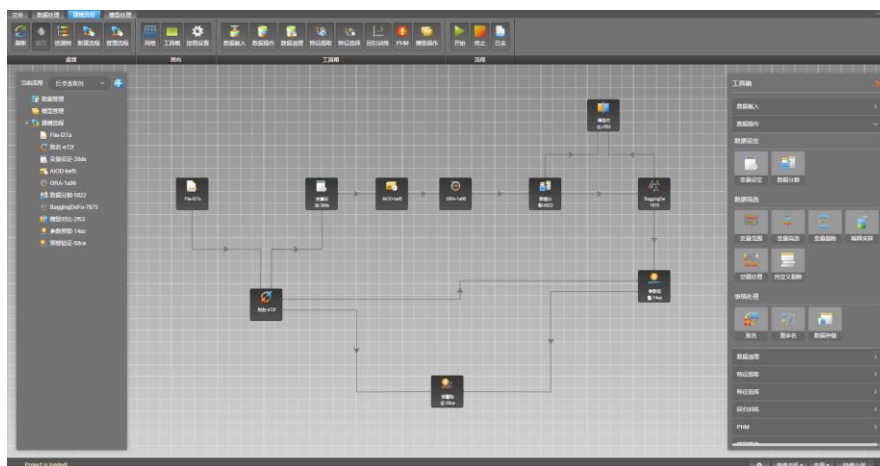


图2 DTEmpower 采用零编码的数据建模流程搭建模式

1.2.3 基于表格视图与交互的数据前处理

DTEmpower 支持在表格视图下，通过低门槛的交互完成数据前处理操作。除覆盖有通用表格的操作与快捷键，还针对性地面向数据分析搭载附加功能，如数据可视化、数据核查与批处理、数据文件拼接等。此外，经过数据前处理完善的数据集可以直接导入建模流程中，完成后续的模式训练等操作。

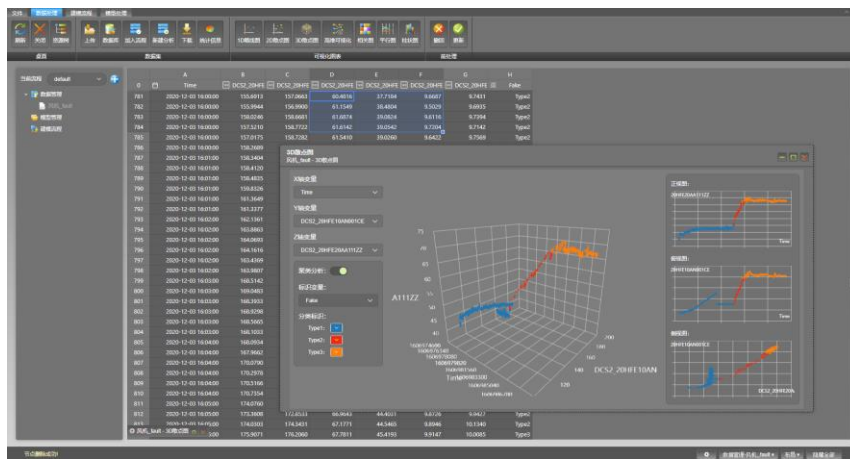


图3 DTEmpower 支持基于表格视图与交互完成数据前处理

1.2.4 智能数据清洗算法

数据中的异常点极大的影响着模型的质量，因此数据清理至关重要。考虑到工业设计数据集的特性，天沓研发了智能数据清洗算法 AIOD。相较于传统的数据清洗算法，它具有以下特点，

(1) 通过自研的智能调度引擎，管控数十种数据清洗算法，包括，Global Outliers Detection、Local Outliers Detection、Contextual Outliers Detection、Regression based OD、Hybrid Auto OD、Cluster Analysis、Classification Analysis 等等，综合考虑数据的整体分布，更精准的挖掘出数据集中的“潜在异常点”；

(2) 使用门槛低，无需任何先验知识，一键完成异常点推荐；

(3) 灵活的异常点剔除规则定义，用户可递进式地观测随着异常点剔除个数的增加而带来的代理模型精度的显著提升，更全面的掌握数据的质量情况。



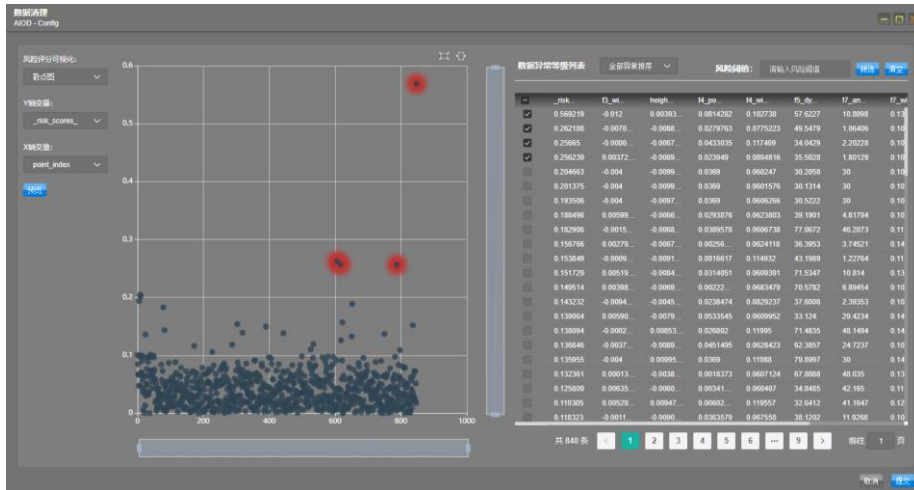


图4 异常点智能识别

1.2.5 针对小规模数据集的 AI Agent 训练

针对工业设计数据的“小数据集”“数据分布不均”等特点，天沓研发了一套智能训练算法 AI Agent，其主要特点包括，

- (1) 使用集成算法提升模型精度和稳定性；
- (2) 通过智能分层分类，级联使用不同置信度来源数据，极小化代理构建成本；
- (3) 通过超参优化，解决参数调节难题，用户无需介入训练过程，一键得到“最优”模型。



图5 船舶兴波阻力数据集 AI Agent 训练同其他算法训练的效果对比



1.2.6 机理模型融合

DTEmpower 支持用户在训练流程中嵌入机理模型，以改善模型的精度和提高模型的可解释性。具体包括，1) 在特征提取环节添加新的自定义特征，2) 在训练环节使用自定义的参数化模型取代黑盒模型，由 DTEmpower 提供优化算法对未知参数进行拟合。DTEmpower 提供模型聚合功能，可以将用户提供的公式模型和数据训练的模型组合，作为单个 DT 模型导出，从而实现数据挖掘与机理模型的融合。

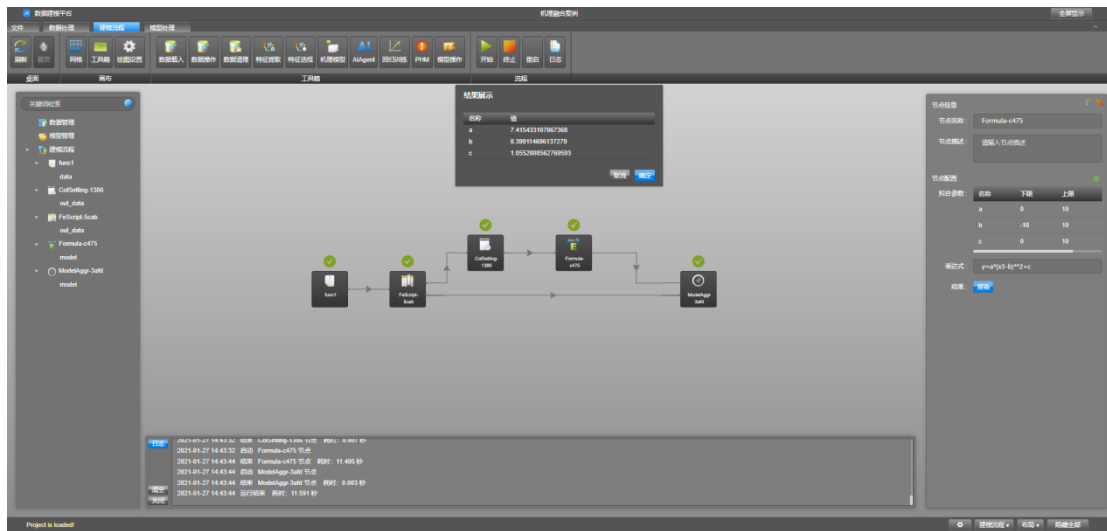


图6 机理模型融合

1.2.7 工业设计场景聚焦——智能优化

DT 模型的一类典型应用场景是，将建立的 DT 模型用于产品的优化设计、设备运行优化等各类优化问题中。DTEmpower 建立的模型可直接导出，同时 AIPOD 支持在计算流程中一键导入 DT 模型，并且可以和脚本、可执行程序进行耦合，实现任意复杂的设计计算流程的搭建。之后便可以借助于 SilverBullet 算法，进行优化问题求解。



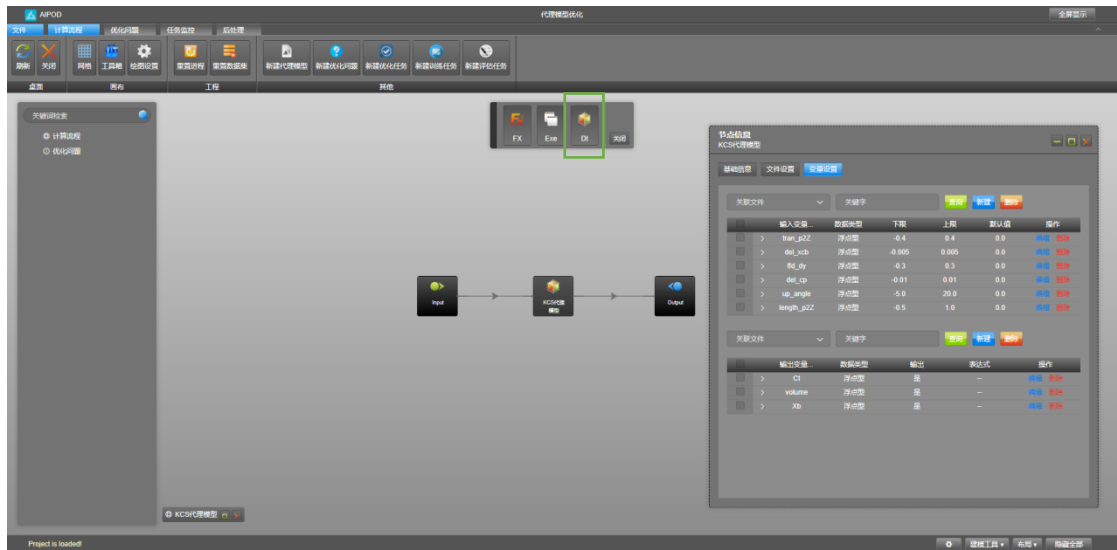


图7 在 AIPOD 中一键导入 DT 模型，耦合建模，进行优化

1.2.8 工业运维场景聚焦——智能监盘

DTEmpower 是一个开放式的数据建模平台，提供了针对不同场景的定制工具箱，针对设备预警场景的智能监盘扩展工具箱正是其中之一。智能监盘扩展工具箱在 DTEmpower 强大的数据建模能力的基础之上，提供了一套针对时间序列的智能预警算法，可以轻松固化专家的判断逻辑，从参数的波动、变化趋势，以及参数和预示模型预测值的相对偏差等多方面对参数进行健康度评估，实现异常早期预警。智能监盘工具箱目前提供有单参数预警、组合参数预警两大控件。

DTEmpower 训练得到的预警模型可以直接导入模型运行模块 DTRun 中，DTRun 可实时接收传感器数据，调用 DT 模型进行数据分析，返回分析结果，实现在线预警。

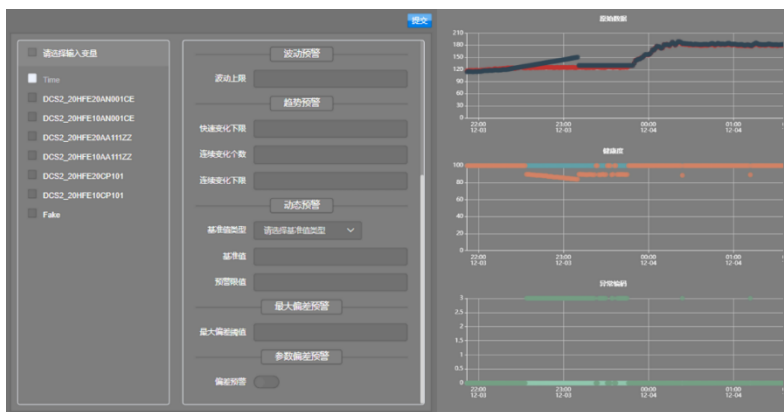


图8 识别时间序列中的异常现象，并给出异常原因，辅助用户快速处置



2 项目组成

2.1 层级结构划分

DTEmpower 项目包含多种不同的层级，如项目、工程、流程等。各层级之间的关联如图 9 所示。

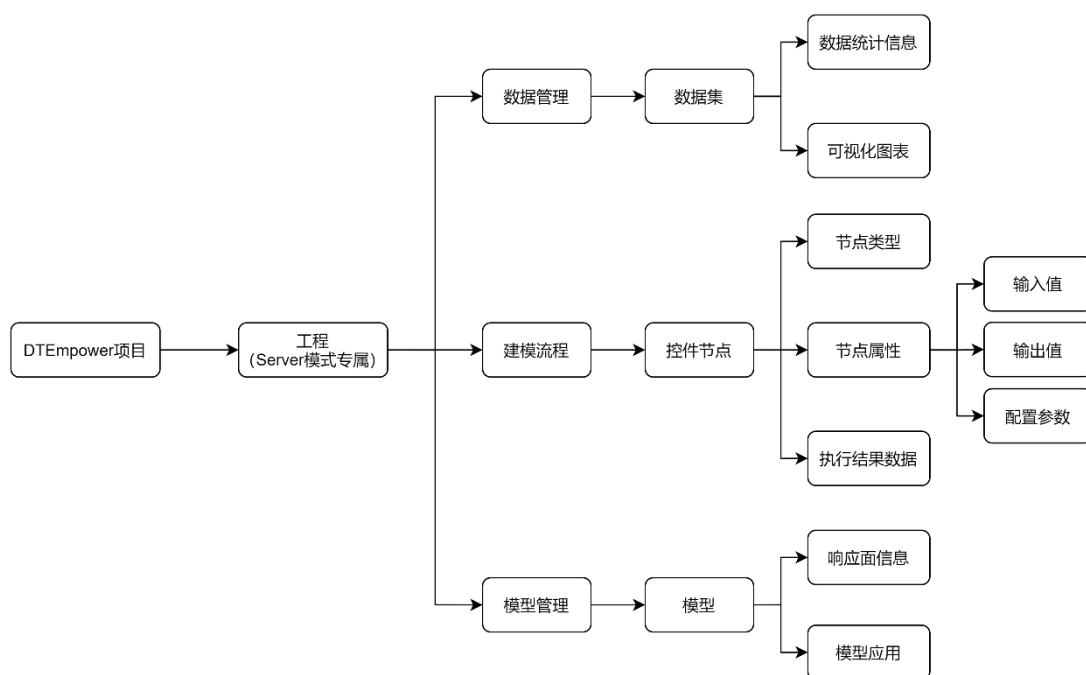


图9 DTEmpower 项目层级

2.2 项目

项目是 DTEmpower 中最高的功能层级。用户每次进入 DTEmpower 时，必须首先选择新建或打开一个项目，之后进行数据分析或建模等工作。

通过双击快捷方式打开 DTEmpower 时，系统将自动新建临时项目。用户可以选择在临时项目中正常完成工作，退出后临时项目将自动删除。用户也可以选择将临时项目保存至本地，后续可以直接打开。

每个项目（包含临时工程）包含项目地址、项目名称、项目简介、项目打开方式与项目端口等，其描述如下：

- 项目地址：对应 DTEmpower 项目文件夹的存储位置；



- 项目名称：标记 DTEmpower 项目名称，与项目存储文件夹名称一致；
- 项目简介：描述 DTEmpower 项目的问题描述、解决方案等；
- 项目打开方式：可选择 DT 客户端方式打开或默认浏览器打开 DTEmpower 项目；
- 项目端口：指定 DTEmpower 项目端口，端口号可在 Ribbon 菜单栏的“文件-设置”中查看。

项目启动界面如图 10 所示：



图10 DTEmpower 项目启动界面

2.3 工程

工程是 Server(服务器)模式下专属的层级结构。其意义在于，当 DTEmpower 软件安装在服务器端时，在同一项目下可能有多个用户同时开展工作。因此为每个用户分配单独的工程，保证其互不干扰且可建立协作。

在本地客户端模式下，工程层级被隐藏，因此用户新建的每个项目中只能含有唯一的工程且不可修改。因此在该语义下，工程与项目的指代含义相同。后文在未特别声明的情况下，工程与项目二者的名称可相互替代。

2.4 数据管理

数据管理用于对当前工程所涉及的原始数据集进行统一管理，并内置丰富的表格操作功能，以支持对数据集的快速修改。同时，数据管理还搭配有包含统计信息在内的数据可视化分析功能。

DTEmpower 工程配置从原始数据集管理(上传)开始。根据应用场景需求，一个工程中可能包含一个或多个原始数据集，如图 11 所示。



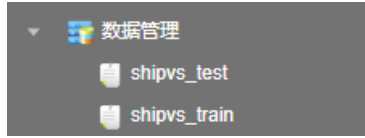
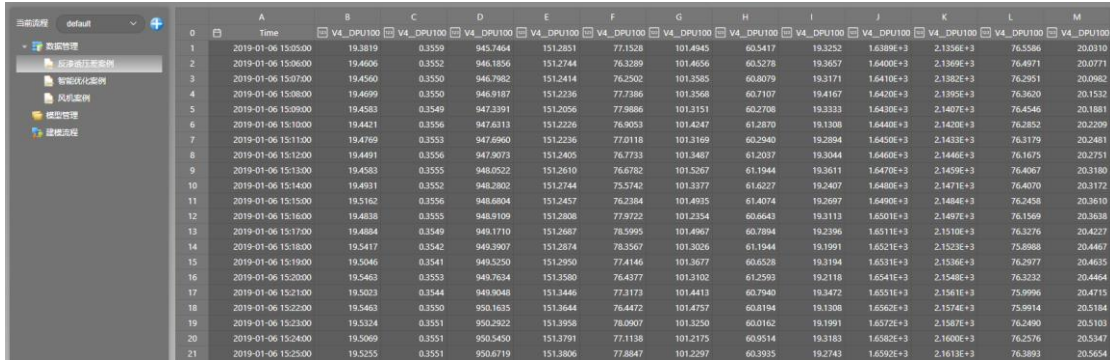


图11 数据集管理

每个原始数据集均以表格的形式向用户呈现，如图 12 所示。



	A	B	C	D	E	F	G	H	I	J	K	L	M
	Time	V4_DPU100	V4_DPU100	V4_DPU100	V4_DPU100	V4_DPU100	V4_DPU100	V4_DPU100	V4_DPU100	V4_DPU100	V4_DPU100	V4_DPU100	V4_DPU100
1	2019-01-06 15:05:00	19.3819	0.3559	945.7464	151.2851	77.1528	101.4645	60.5417	19.3252	1.6389E+3	2.1356E+3	76.5586	20.0310
2	2019-01-06 15:06:00	19.4606	0.3552	946.1856	151.2744	76.3289	101.4656	60.5278	19.3657	1.6400E+3	2.1369E+3	76.4971	20.0771
3	2019-01-06 15:07:00	19.4560	0.3550	946.7982	151.2414	76.2502	101.3585	60.8079	19.3171	1.6410E+3	2.1382E+3	76.2951	20.0982
4	2019-01-06 15:08:00	19.4699	0.3550	946.9187	151.2236	77.7386	101.3568	60.7107	19.4167	1.6420E+3	2.1395E+3	76.3620	20.1532
5	2019-01-06 15:09:00	19.4583	0.3549	947.3391	151.2056	77.9886	101.3151	60.2708	19.3333	1.6430E+3	2.1407E+3	76.4546	20.1881
6	2019-01-06 15:10:00	19.4421	0.3556	947.6313	151.2226	76.9053	101.4247	61.2870	19.1308	1.6440E+3	2.1420E+3	76.2852	20.2209
7	2019-01-06 15:11:00	19.4769	0.3553	947.6960	151.2236	77.0118	101.3169	60.2940	19.2894	1.6450E+3	2.1433E+3	76.3179	20.2481
8	2019-01-06 15:12:00	19.4491	0.3556	947.9073	151.2405	76.7733	101.3487	61.2037	19.3044	1.6460E+3	2.1446E+3	76.1675	20.2751
9	2019-01-06 15:13:00	19.4583	0.3555	948.0522	151.2610	76.6782	101.5267	61.1944	19.3611	1.6470E+3	2.1459E+3	76.4067	20.3180
10	2019-01-06 15:14:00	19.4931	0.3552	948.2802	151.2744	75.5742	101.3377	61.6227	19.2407	1.6480E+3	2.1471E+3	76.4070	20.3172
11	2019-01-06 15:15:00	19.5162	0.3556	948.6804	151.2457	76.2384	101.4935	61.4074	19.2697	1.6490E+3	2.1484E+3	76.2458	20.3610
12	2019-01-06 15:16:00	19.4838	0.3555	948.9109	151.2008	77.9722	101.2354	60.6643	19.3113	1.6501E+3	2.1497E+3	76.1569	20.3698
13	2019-01-06 15:17:00	19.4884	0.3549	949.1710	151.2687	78.5995	101.4967	60.7894	19.2396	1.6511E+3	2.1510E+3	76.3276	20.4227
14	2019-01-06 15:18:00	19.5417	0.3542	949.3907	151.2874	78.3567	101.3036	61.1944	19.1991	1.6521E+3	2.1523E+3	75.9888	20.4467
15	2019-01-06 15:19:00	19.5046	0.3541	949.5250	151.2950	77.4146	101.3677	60.6528	19.3194	1.6531E+3	2.1536E+3	76.2977	20.4655
16	2019-01-06 15:20:00	19.5463	0.3553	949.7634	151.3580	76.4377	101.3102	61.2592	19.2116	1.6541E+3	2.1549E+3	76.3232	20.4464
17	2019-01-06 15:21:00	19.5023	0.3544	949.9048	151.3446	77.3173	101.4413	60.7940	19.3472	1.6551E+3	2.1561E+3	75.9996	20.4715
18	2019-01-06 15:22:00	19.5463	0.3550	950.1635	151.3644	76.4472	101.4757	60.8194	19.1308	1.6562E+3	2.1574E+3	75.9914	20.5184
19	2019-01-06 15:23:00	19.5334	0.3551	950.2932	151.3958	76.9097	101.3250	60.0162	19.1991	1.6572E+3	2.1587E+3	76.2490	20.5193
20	2019-01-06 15:24:00	19.5069	0.3551	950.5450	151.3791	77.1138	101.2175	60.9514	19.3183	1.6582E+3	2.1600E+3	76.2576	20.5347
21	2019-01-06 15:25:00	19.5255	0.3551	950.6719	151.3806	77.8847	101.2297	60.3935	19.2743	1.6592E+3	2.1613E+3	76.3893	20.5694

图12 原始数据集表格

在表格查看界面中，还内置有丰富的表格操作功能，用以辅助用户快速调整数据集，如图 13 所示。完成处理的数据集可以无缝衔接到后续建模流程操作。

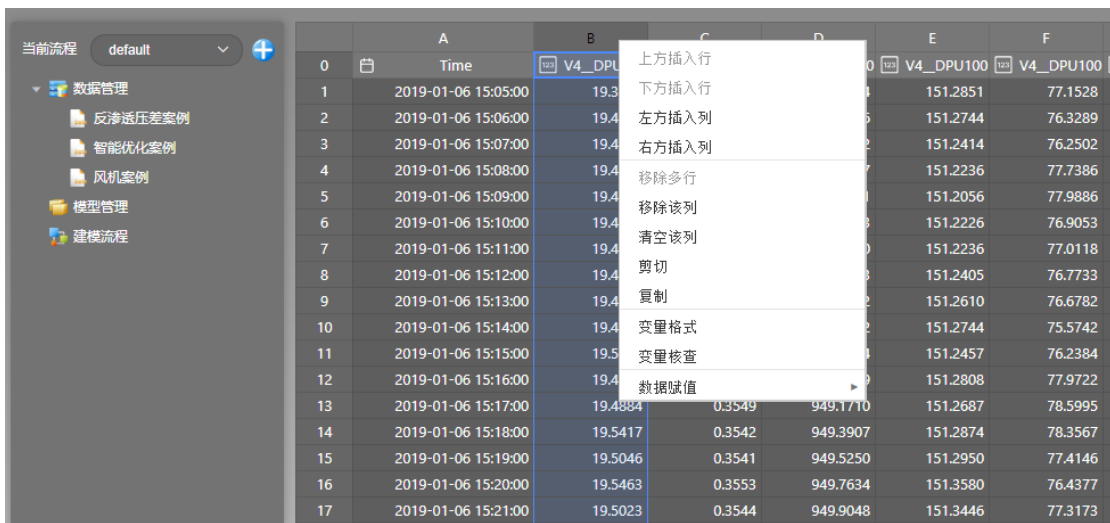


图13 内置表格操作功能

针对每一个原始数据集，DTEmpower 提供各个参数的统计信息如图 14 所示。



变量名	最小值	最大值
Time	1546787100	1548925140
V4_DPU1001_HW_Ai020201_PV	16.12495804	21.43753052
V4_DPU1001_SH0466_AALM5_PV	0.298329473	0.385533184
V4_DPU1001_SH0460_AALM01_PV	876.8848877	1545.877686
V4_DPU1001_SH0460_AALM03_PV	129.4109497	164.1344757
V4_DPU1001_HW_Ai020104_PV	68.20806122	80.12266541
V4_DPU1001_SH0466_AALM6_PV	95.00235748	113.1315918
V4_DPU1001_HW_Ai020106_PV	51.47693634	61.62270355
V4_DPU1001_HW_Ai020107_PV	16.59487915	20.92708778
V4_DPU1001_SH0428_AALM4_PV	1638.935547	13424.61426
V4_DPU1001_SH0428_AALM01_PV	2135.623535	17674.97266
V4_DPU1001_SH0466_AALM7_PV	67.71627045	81.65318298
V4_DPU1001_SH0460_AALM04_PV	20.00124359	32.99402237

行数: 8856 列数: 13

图14 原始数据集统计信息

此外，针对每个原始数据集，DTEmpower 还提供了包含 1D 箱线图、2D 散点图、3D 散点图、高维可视化、相关图、平行图和柱状在内的数据可视化图表，用以辅助用户对原始数据进行快速分析并建立直观的认识。

2.5 数据源管理

数据源管理为用户提供了快速连接至数据库并获取数据的功能。对于企业用户而言，数据的来源通常并不是离散的数据文件，而是从数据库中批量获取的。在数据源管理中保存了用户从数据库中提取数据的配置操作，方便用户日常分析数据。如图 15 中所示为空白数据源的配置界面，用户在首次连接过程中手动填写服务器 IP、用户名、数据表名等基本信息并保存，后续即可在数据源管理页面快速复用，如图 16 所示。





图15 数据源配置界面

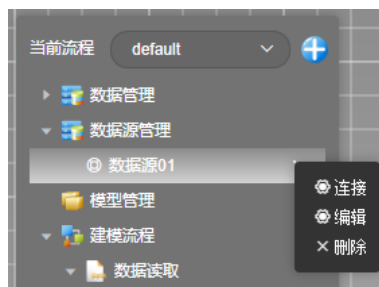


图16 通过数据源管理快速连接至数据库

需要注意的是，图 16 中所显示的数据源管理层级是默认隐藏的，当用户手动配置并保存了一个数据源后，才会在资源树中保持显示。

2.6 建模流程

建模流程即数据挖掘过程，通过建模流程的配置与执行能够实现数据到模型的转换。其一般包含原始数据读取、数据清洗、特征工程、模型训练（以及模型存储）几个部分。DTEmpower 提供图形化的建模流程搭建功能，如图 17 所示。



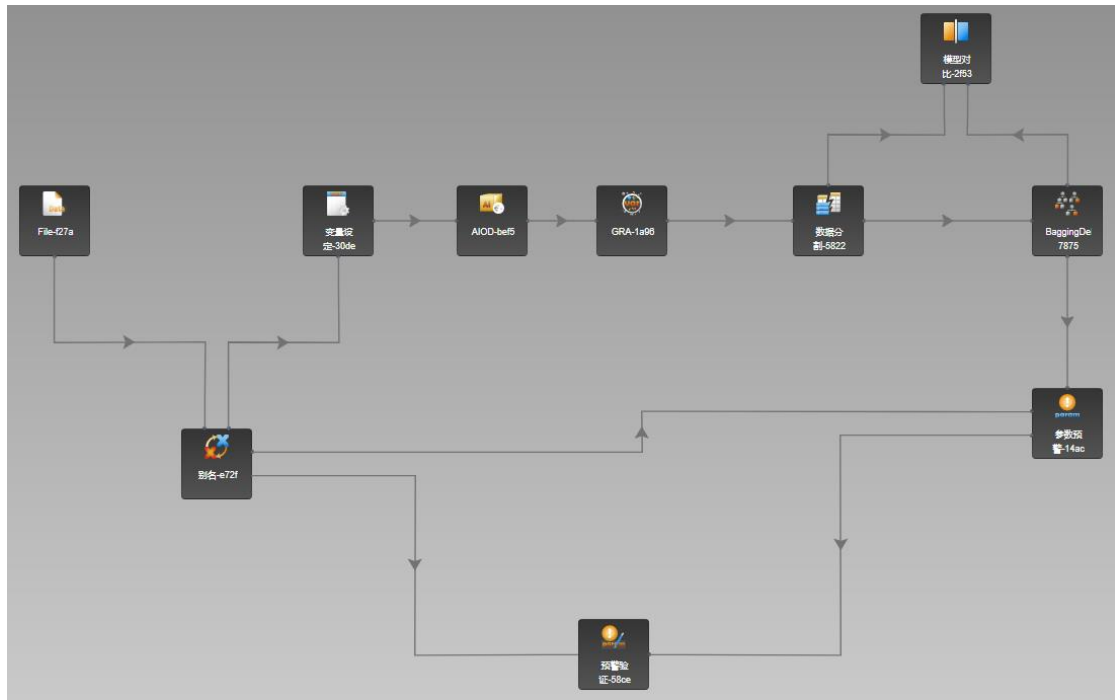


图17 建模流程

原始数据集处理完成后，即可将其加入至建模流程，作为数据挖掘流程的起点，通过与其它功能节点的拖拽与连接即可实现完整的数据挖掘过程，实现数据到知识（模型）的转化。

为适应数据挖掘的需要，在同一工程下，可同时建立多个建模流程并进行流程间的切换，且各个建模流程共享相同的数据管理以及模型管理资源，如图 18 所示。

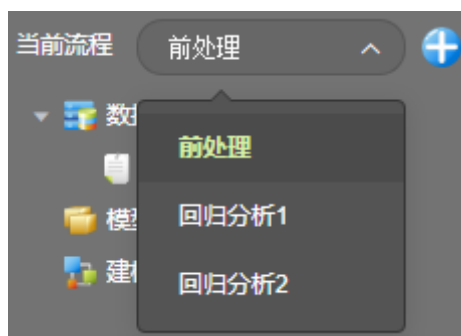


图18 多个建模流程间切换

2.6.1 节点

节点为建模流程的基本组成单元，每个节点提供一种功能（或方法）的实现，DTEmpower 提供了大量的功能节点，如图 19 所示。



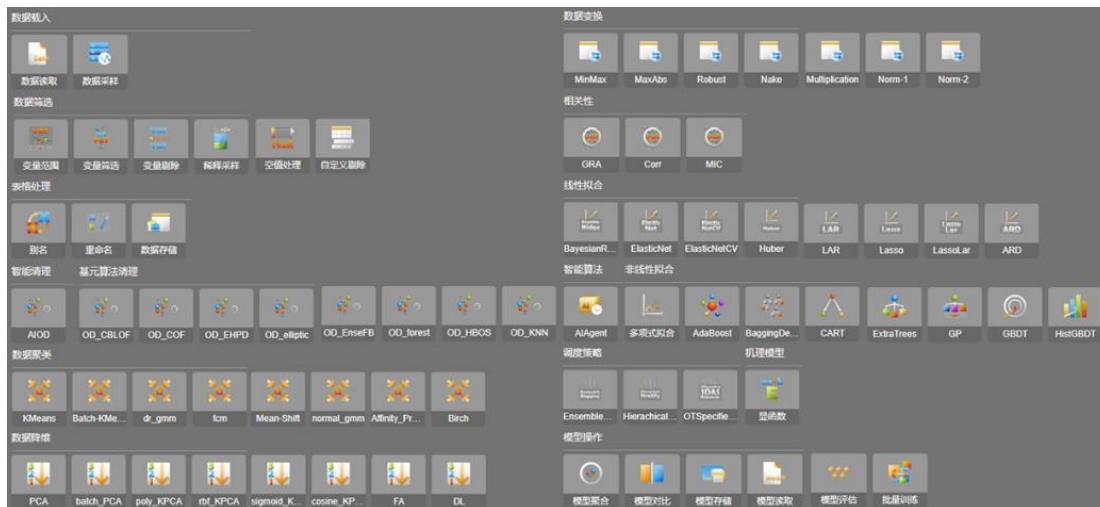


图19 功能节点

执行建模流程时，DTEmpower 自动根据建模流程中节点的连接顺序依次执行各节点，最终实现数据到知识（模型）的转化。

此外，作为开放式的数据建模平台，DTEmpower 还针对不同场景提供了定制化的功能节点，如针对设备运行早期预警场景的智能监盘扩展工具箱，如图 20 所示。



图20 智能监盘工具箱

2.6.1.1 节点类型

根据节点实现功能（或方法）不同，DTEmpower 内将节点分为数据载入、数据操作、数据清理、特征生成、敏感性分析、回归训练、智能监盘和模型操作 8 类，如图 21 所示。



图21 节点类型

不同类型节点的功能及节点详情将在手册第 5 章中进行展开描述。



2.6.1.2 节点属性

1) 输入值

除数据载入和模型读取节点外，所有节点的执行都依托于相应的输入值，若输入值缺失，将导致该节点执行失败。此处以 AIAgent 节点为例，AIAgent 为模型训练节点，因此其输入值为指定了输入变量和输出变量的训练数据集。

2) 输出值

所有节点执行完成后，都会产生相应的输出值，对于非流程末端的节点而言，其输出至通常作为下游节点的输入值进行传递。此处以 AIAgent 节点为例，AIAgent 为模型训练节点，因此其输出值为训练得到的模型。

3) 配置参数

节点执行时，除了需要依托于输入之外，通常还需要相关的节点参数进行配置，不同的配置参数可能获得不同的输出值，鼠标左键单击节点即可查看和编辑节点的配置参数，如图 22 所示。



图22 节点配置参数

此处以多项式拟合节点为例，如上图所示。多项式拟合节点除了节点名称、节点描述、生成模型名称外，还包含多项式拟合时的最高阶数设定及其他参数配



置选项。不同的节点拥有不同的节点配置参数，各节点配置参数的详情将在第 5 章展开进行介绍。

2.6.1.3 执行结果数据

节点执行得到的输出值，通常可以作为节点执行结果数据供用户查看和分析。当节点执行完成后，可通过在工程树状资源菜单中单击对应节点的方式展开节点执行结果数据，如图 23 所示。

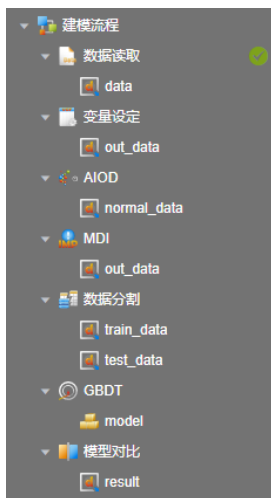


图23 节点执行结果数据

单击任意执行结果数据，即可对该数据进行查看，通常而言，不同节点的执行结果数据不同，以数据清理节点 AIOD 为例，其执行结果为剔除异常数据之后的数据列表，如图 24 所示。不同节点的执行结果数据不同，各节点的执行结果数据将在第五章展开进行介绍。

	A	B	C	D	E	F	G	H	I	J	K	L
0	V4_DPU100	V4_DPU100	V4_DPU100	V4_DPU100	V4_DPU100	V4_DPU100	V4_DPU100	V4_DPU100	V4_DPU100	V4_DPU100	V4_DPU100	V4_DPU100
1	19.3819	76.5586	0.3559	945.7464	151.2851	77.1528	101.4945	60.5417	19.3252	1.6389E+3	2.1356E+3	20.0310
2	19.4606	76.4971	0.3552	946.1856	151.2744	76.3289	101.4656	60.5278	19.3657	1.6400E+3	2.1369E+3	20.0771
3	19.4560	76.2951	0.3550	946.7982	151.2414	76.2502	101.3585	60.8079	19.3171	1.6410E+3	2.1382E+3	20.0982
4	19.4699	76.3620	0.3550	946.9187	151.2236	77.7386	101.3568	60.7107	19.4167	1.6420E+3	2.1395E+3	20.1532
5	19.4583	76.4546	0.3549	947.3391	151.2056	77.9886	101.3151	60.2708	19.3333	1.6430E+3	2.1407E+3	20.1881
6	19.4421	76.2852	0.3556	947.6313	151.2226	76.9053	101.4247	61.2870	19.1308	1.6440E+3	2.1432E+3	20.2209
7	19.4769	76.3179	0.3553	947.6960	151.2236	77.0118	101.3169	60.2940	19.2894	1.6450E+3	2.1433E+3	20.2481
8	19.4491	76.1675	0.3556	947.9073	151.2405	76.7733	101.3487	61.2037	19.3044	1.6460E+3	2.1446E+3	20.2751
9	19.4583	76.4067	0.3555	948.0522	151.2610	76.6782	101.5267	61.1944	19.3611	1.6470E+3	2.1459E+3	20.3180
10	19.4931	76.4070	0.3552	948.2882	151.2744	75.5742	101.3377	61.6227	19.2407	1.6480E+3	2.1471E+3	20.3172
11	19.5162	76.2458	0.3556	948.6804	151.2457	76.2384	101.4935	61.4074	19.2697	1.6490E+3	2.1484E+3	20.3610
12	19.4838	76.1569	0.3555	948.9189	151.2808	77.9722	101.2354	60.6643	19.3113	1.6501E+3	2.1497E+3	20.3638
13	19.4884	76.3276	0.3549	949.1710	151.2687	78.5995	101.4967	60.7894	19.2396	1.6511E+3	2.1510E+3	20.4227
14	19.5417	75.8988	0.3542	949.3907	151.2874	78.3567	101.3026	61.1944	19.1991	1.6521E+3	2.1523E+3	20.4467
15	19.5046	76.2977	0.3541	949.5250	151.2950	77.4146	101.3677	60.6528	19.3194	1.6531E+3	2.1536E+3	20.4635
16	19.5463	76.3232	0.3553	949.7634	151.3580	76.4377	101.3102	61.2593	19.2118	1.6541E+3	2.1548E+3	20.4464
17	19.5023	75.9996	0.3544	949.9048	151.3446	77.3173	101.4413	60.7940	19.3472	1.6551E+3	2.1561E+3	20.4715
18	19.5463	75.9914	0.3550	950.1635	151.3644	76.4472	101.4757	60.8194	19.1308	1.6562E+3	2.1574E+3	20.5184
19	19.5324	76.2490	0.3551	950.2922	151.3958	78.0907	101.3250	60.0162	19.1991	1.6572E+3	2.1587E+3	20.5103
20	19.5069	76.2576	0.3551	950.5450	151.3791	77.1138	101.2175	60.9514	19.3183	1.6582E+3	2.1600E+3	20.5347
21	19.5255	76.3893	0.3551	950.6719	151.3806	77.8847	101.2297	60.3935	19.2743	1.6592E+3	2.1613E+3	20.5654
22	19.5069	76.1234	0.3528	950.8403	151.3565	76.4078	101.2094	61.4653	19.2986	1.6602E+3	2.1625E+3	20.5799
23	19.5509	73.7615	0.3567	951.3540	151.3445	74.8985	97.3294	58.7824	18.9838	1.6612E+3	2.1638E+3	20.6760
24	19.5718	74.3874	0.3556	951.2114	151.3669	74.6554	97.5011	59.2986	19.0394	1.6622E+3	2.1650E+3	20.9596

图24 节点执行结果数据



2.6.2 节点连接

节点间的连接表征数据的传递关系，节点连接后，系统会自动对上游节点输出值和下游节点输入值进行匹配，如图 25 所示。



图25 节点连接

空值处理流程如上图所示。原始数据首先通过数据读取节点载入，原始数据即作为数据读取节点的输出值；然后，原始数据传递给空值处理节点作为空值处理节点的输入值，经空值处理节点处理后的数据作为空值处理节点的输出值。根据上述描述，数据读取节点与空值处理节点存在先后的执行顺序，且两个节点间存在数据的传递，因此需要顺序连接两个节点，节点连接后，系统会自动对输入值和输出值进行关联，需要注意的是，连线配置中的输入值和输出值是相对于连线本身而言的。

2.7 模型管理

模型管理用于对当前工程获得的模型进行统一管理，支持模型的存储复用。数据建模平台工程截止于模型（知识）的管理，根据应用场景需求，一个工程下可能存在一个至多个模型，用户可将这些模型存储至工程资源树中，如图 26 所示。



图26 模型管理



模型管理下的每个模型都支持下载到本地磁盘中，以便实现跨平台的智能工业设计或智能工业运维。

2.7.1 模型响应面信息

单击模型管理中的指定模型，即可查看该模型的响应面信息如图 27 所示。

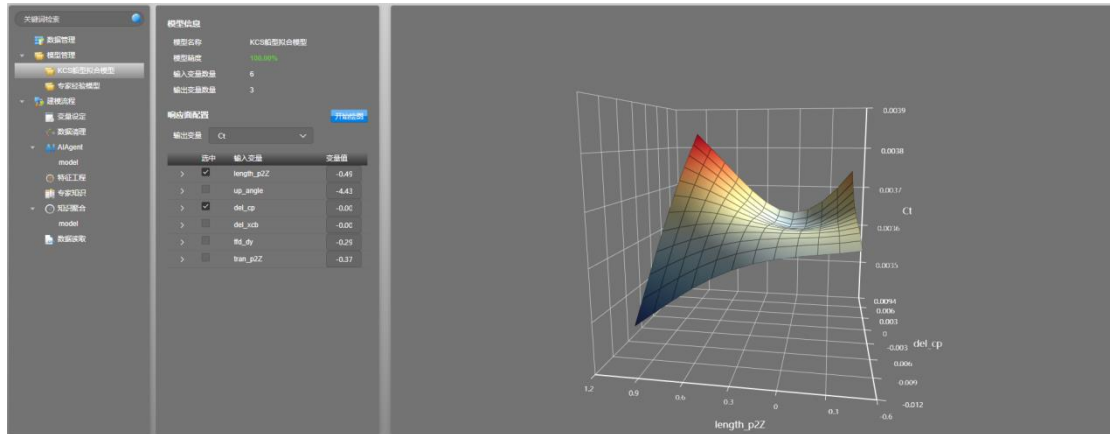


图27 代理模型响应面

在查看模型响应面信息时，最多支持用户指定两个输入变量作为自变量，一个输出变量作为因变量进行响应面的绘制，需要注意的时，绘制该响应面时，其它未被指定为自变量的输入变量将保持各自的变量值不变。

2.7.2 模型应用

DTEmpower 支持直接使用模型管理中的模型，对于给定的输入变量值，快速评估得到输出变量值，如图 28 所示。



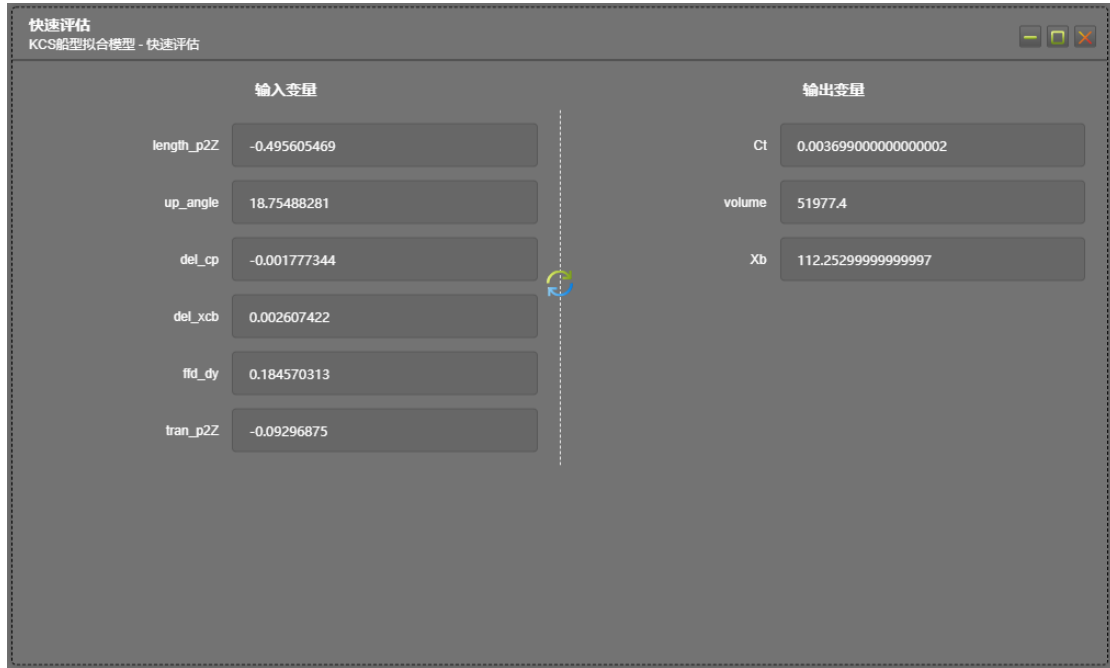


图28 快速评估

3 界面布局

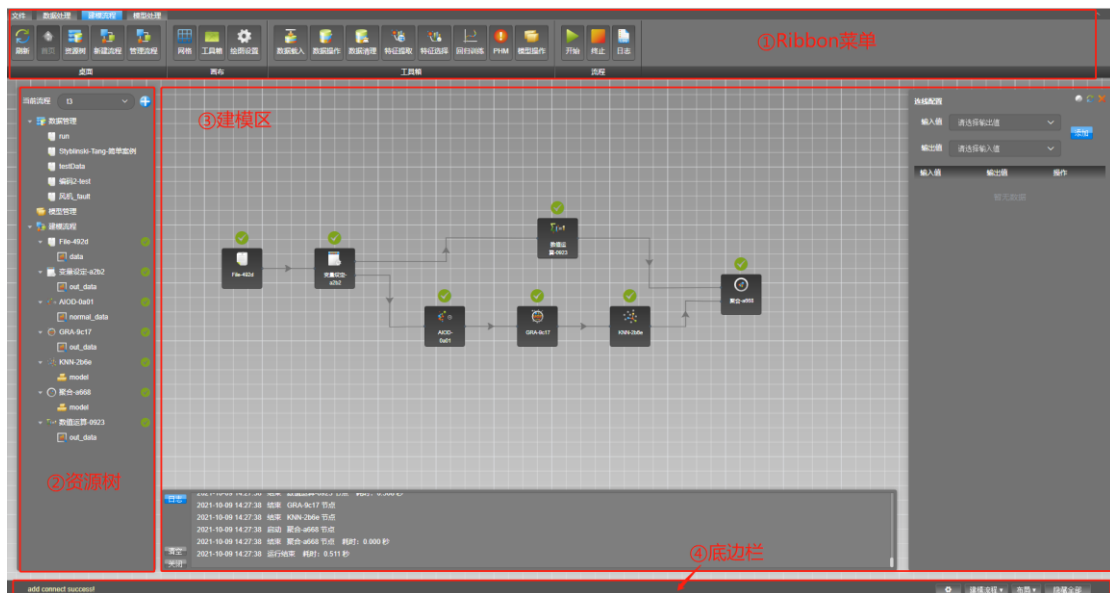


图29 工程详情页面



如图 29 所示，工程详情页面分为四个部分，分别为：顶部的 Ribbon 菜单栏，左侧资源树，右侧建模区，以及底边栏。Ribbon 菜单和桌面跟随用户选中的资源进行同步切换。

3.1 Ribbon 菜单

Ribbon 菜单中包含“文件”、“数据处理”、“建模流程”、“模型处理”4 个基础选项。

3.1.1 文件



图30 文件菜单

文件菜单选项如图 30 所示，其按钮功能描述如下：

- 刷新：刷新当前桌面内容；
- 设置：查看当前项目的 URL 地址；
- 皮肤：切换界面风格，支持 6 种不同主题颜色的切换；
- 语言：切换软件语言，支持简体中文与英文二者间切换；
- 帮助：打开 DTEmpower 产品说明页面；
- 新建：创建新的项目；
- 打开：打开已创建的项目；
- 保存：保存当前项目。

（注：上述新建、打开与保存三个按钮功能仅在本地客户端运行模式下有效，见本手册 2.2 节“项目”）

3.1.2 数据处理

数据处理菜单中的内容随不同工作阶段有所区别。在画布状态下，数据处理仅作为数据上传的窗口之一，大部分功能处于关闭状态。





图31 数据处理菜单

选择左侧资源树-数据管理中的数据文件，可切换至数据表格查看状态，并激活数据处理中的统计与绘图等基础功能。数据处理菜单选项如图 31 所示，其按钮功能描述如下：

- 刷新：刷新桌面信息；
- 关闭：关闭当前桌面，回到建模流程画布；
- 资源树：显示或隐藏左侧资源树；
- 上传：上传本地数据文件至数据管理；
- 数据库：连接数据库，提取数据文件至数据管理；
- 加入流程：以数据读入节点的方式将当前数据集加入至当前建模流程中；
- 新建分析：以数据读入节点的方式将当前数据集新建一个同名流程（默认进入表格式建模流程）；
- 下载：下载 DTable 格式（DTEmpower 专用数据格式）数据至本地磁盘；
- 统计信息：查看当前数据集数据统计信息；
- 1D 箱线图：选择变量绘制 1D 箱线图；
- 2D 散点图：选择变量绘制 2D 散点图；
- 3D 散点图：选择变量绘制 3D 散点图；
- 高维可视化：选择变量绘制高维可视化视图；
- 相关图：选择变量绘制相关图；
- 平行图：选择变量绘制平行图；
- 柱状图：选择变量绘制柱状图；
- 折线图：选择变量绘制折线图；
- 加载全部：开启表格编辑模式；
- 撤回：在表格编辑模式下，撤回对当前数据的修改；
- 更新：在表格编辑模式下，更新对当前数据的修改。



当进入表格式建模流程时，数据处理菜单中将激活更多的功能，如图 32 中红框内所示：



图32 表格式建模流程下的数据处理菜单

新激活的按钮功能描述如下：

- 数据操作：表格式建模工具箱内容切换至数据操作功能节点列表；
- 数据清理：表格式建模工具箱内容切换至数据清理功能节点列表；
- 特征提取：表格式建模工具箱内容切换至特征提取功能节点列表；
- 特征选择：表格式建模工具箱内容切换至特征选择功能节点列表；
- 回归训练：表格式建模工具箱内容切换至回归训练功能节点列表；
- 开始：执行当前表格式建模流程；
- 终止：终止当前表格式建模流程执行；
- 日志：显示或隐藏当前流程运行日志信息；
- 节点删除：删除资源树中选中节点及其后所有节点。

3.1.3 建模流程

在画布状态下，可激活建模流程中的全部按钮功能，如图 33 所示：



图33 建模流程菜单

- 刷新：刷新当前桌面信息；
- 首页：返回建模画布；
- 资源树：显示或隐藏左侧资源树；
- 新建流程：打开新建流程窗口；
- 管理流程：打开管理流程窗口；
- 网格：显示或隐藏话务网格；
- 工具箱：显示或隐藏节点工具箱；
- 绘图设置：建模流程显示配置；



- 数据载入：拖拽式建模工具箱内容切换至数据载入功能节点列表；
- 数据操作：拖拽式建模工具箱内容切换至数据操作功能节点列表；
- 数据清理：拖拽式建模工具箱内容切换至数据清理功能节点列表；
- 特征生成：拖拽式建模工具箱内容切换至特征生成功能节点列表；
- 敏感性分析：拖拽式建模工具箱内容切换至敏感性分析功能节点列表；
- 回归训练：拖拽式建模工具箱内容切换至回归训练方法功能节点列表；
- 时序预测：拖拽式建模工具箱内容切换至时序预测方法功能节点列表；
- 智能监盘：拖拽式建模工具箱内容切换至智能监盘功能节点列表；
- 模型操作：拖拽式建模工具箱内容切换至模型操作功能节点列表；
- 开始：执行当前建模流程；
- 终止：终止当前建模流程执行；
- 日志：显示或隐藏当前流程运行日志信息。

3.1.4 模型处理



图34 模型处理菜单

选择左侧资源树下的模型文件，可激活模型处理中的功能。模型处理菜单如图 34 所示，其按钮功能描述如下：

- 刷新：刷新当前桌面信息；
- 关闭：关闭当前桌面，返回建模流程画布；
- 资源树：显示或隐藏左侧资源树；
- 模型详情：模型变量、精度、响应面信息查看；
- 误差数据：训练数据及预测偏差数据查看；
- 精度数据：模型精度指标结果查看；
- 下载：下载当前模型至本地磁盘；
- 保存：保存当前模型至当前工程模型管理列表中；



- 评估：给定输入参数，利用模型进行快速计算；
- 线性拟合：模型拟合曲线查看；
- 误差对比：真实值&预测值折线查看。

注意：对于不同类型的模型，如回归模型、时序模型、智能监盘模型等，模型处理界面存在一定的差异。本节中内容以回归模型为例。

3.2 资源树

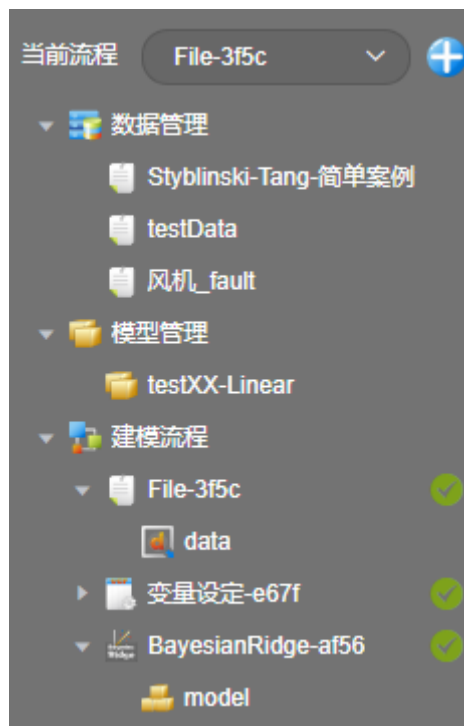


图35 工程资源树

工程资源树为当前工程的全部信息，如图 35 所示。共分为流程切换以及数据管理、模型管理和建模流程三个主菜单，其各自功能描述如下：

- 流程切换：显示当前建模流程，通过下拉菜单切换流程，以及新建流程等功能；
- 数据管理菜单：管理该工程下所有的数据集，来源于本地磁盘上传原始数据集，或建模流程中生成的新数据集。当前工程下的所有流程共享相同的数据管理内容。



- **模型管理菜单：**管理该工程下所有的模型，来源于本地磁盘上传的“.dt”格式模型文件，或建模流程中生成的新模型（知识）。当前工程下的所有流程共享相同的模型管理内容。
- **建模流程菜单：**下包含流程中所有的功能节点，当某节点执行完成后，节点作为二级菜单支持展开，其展开内容为执行结果信息。

3.3 建模区

建模区具有两种不同的显示状态：表格式建模区与拖拽式建模区，分别对应两种建模流程。

3.3.1 表格式建模区

表格式建模区的形态如图 36 所示，主要包含左侧的表格查看区以及右侧的数据处理工具箱。

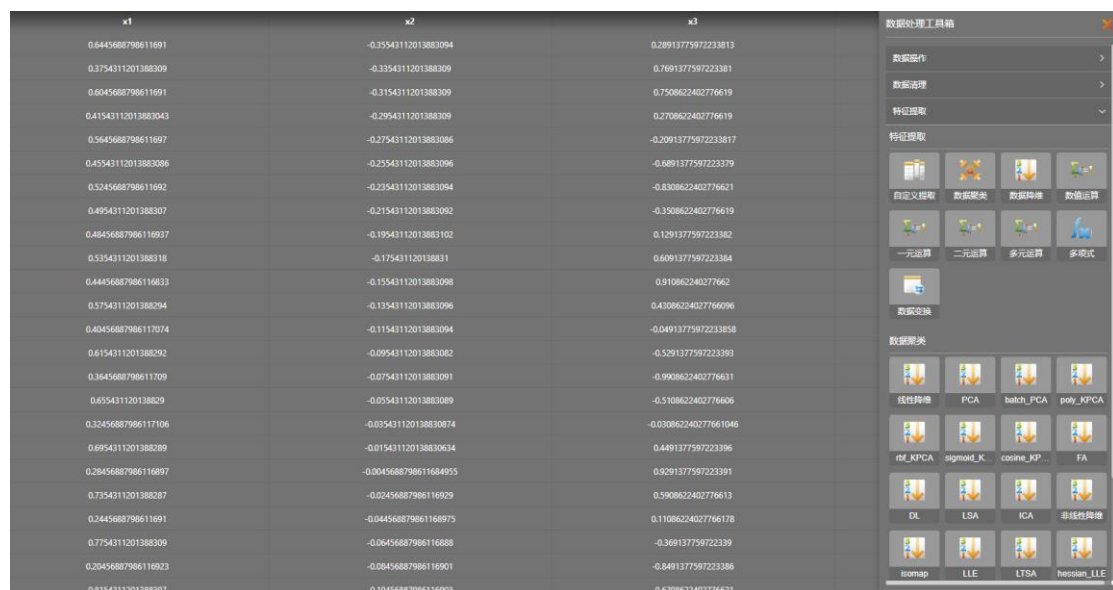


图36 表格式建模区

- **表格查看区：**以表格形式展示当前数据集，提供对每次数据处理操作的直观认识。与资源树中数据管理下的表格查看功能不同，建模流程中的表格查看区不能对数据直接进行修改。
- **数据处理工具箱：**由顶部 Ribbon 菜单中的数据处理的功​​能唤出，单击工具箱内节点可对当前数据完成操作，操作结果在表格查看区内更新。应注



意的是，在表格式建模方式下，只有最新的数据（data）支持唤出数据处理工具箱。若需回退至某节点，需进行“节点删除”操作。

3.3.2 拖拽式建模区

表格式建模区的形态如图 37 所示，主要包含左侧的建模画布区以及右侧的工具箱和节点面板。

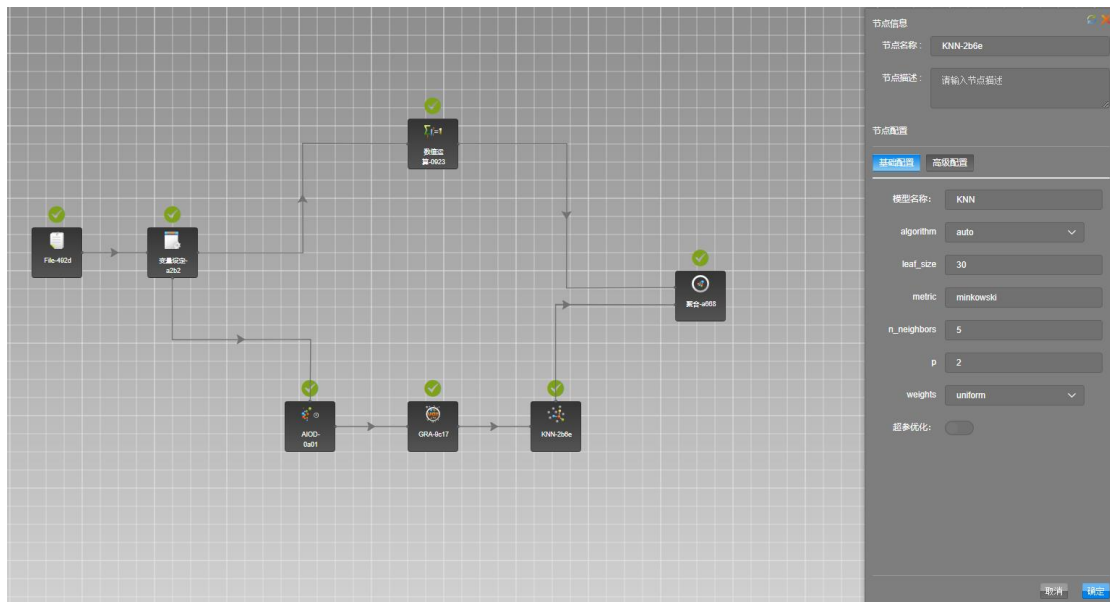


图37 拖拽式建模区

- 建模画布区：工具箱中的功能控件通过拖拽的方式加入建模画布区中，并添加连接线构建完整的建模流程。
- 工具箱：由顶部 Ribbon 菜单唤出，用于向建模画布区添加功能节点。
- 节点面板：点击画布中的节点或连接线唤出，用于展示节点信息或修改节点配置。

3.4 底边栏

底边栏包括左侧的运行信息提示以及右侧的四个功能按钮，其功能描述分别如下：

- 运行信息提示：提示当前流程的运行状态，如添加节点、删除连接线、修改绘图设置等；
- 桌面清空：一键清空当前流程下的全部展开页面，并返回建模画布；



- 桌面切换：在各个展开页面之间切换，如图 38 所示；
- 修改布局：显示或隐藏日志、资源树、工具箱等；
- 隐藏/显示全部：日志、资源树、工具箱等全部显示或隐藏。



图38 桌面切换

4 基础操作说明

4.1 软件安装

4.1.1 安装启动程序

安装 DTEmpower 前，首先需要明确计划安装路径与当前用户权限。

如果计划安装在系统盘，则需要对 exe 文件上单击鼠标右键选择【以管理员身份运行】安装程序如图 39 所示。

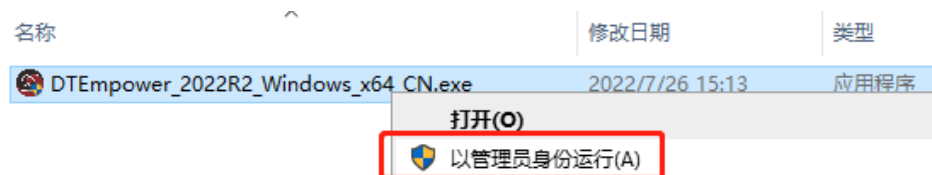


图39 安装在系统盘需以管理员身份运行

进入安装向导如图 40 所示：



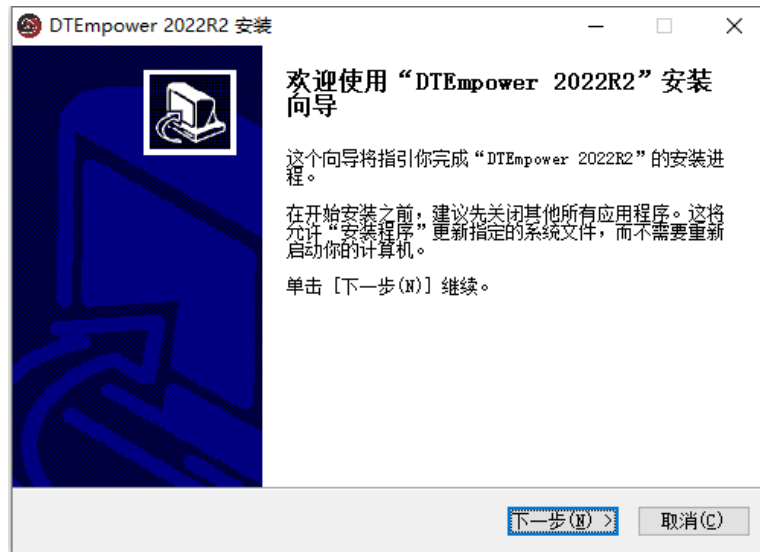


图40 DTEmpower 安装向导

4.1.2 浏览安装路径

在安装向导中单击【下一步】按钮，通过浏览按钮选择安装路径，如图 41 所示。

注意：选择安装路径需为全英文。

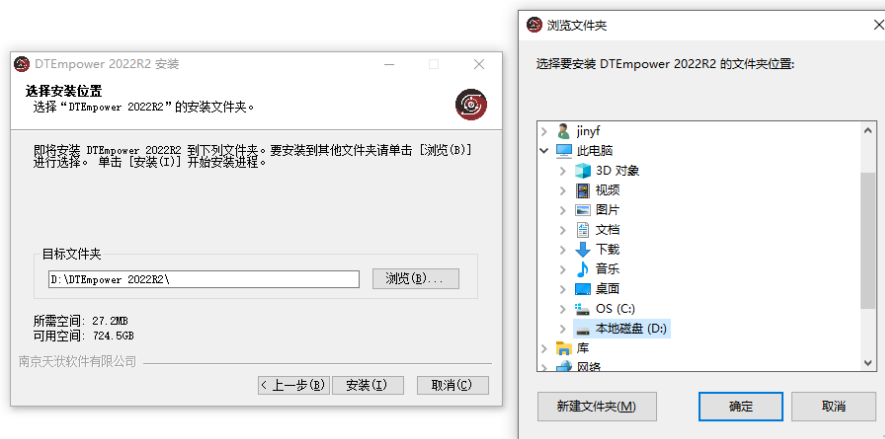


图41 选择安装路径

4.1.3 文件解压与拷贝

点击【安装】按钮，即可开始进行安装，如图 42 所示。提示安装完成后，单击【完成】按钮即可结束安装。



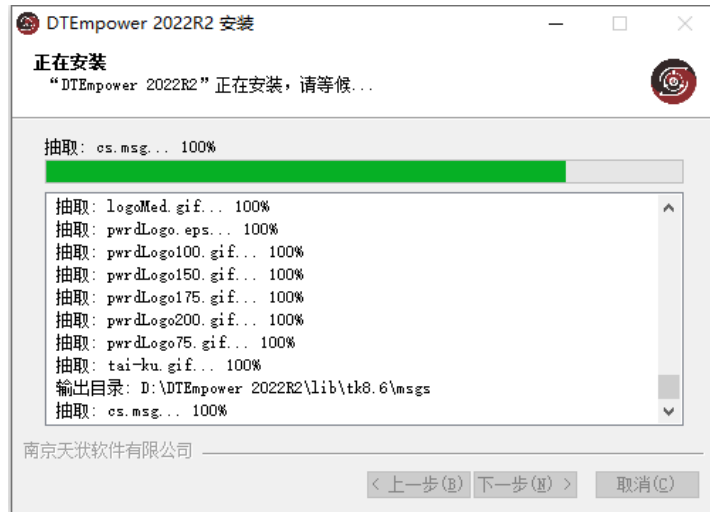


图42 文件解压与拷贝

4.1.4 以管理员身份运行程序

软件安装时，若将安装路径指向了需要管理员权限的目录（如C:\Program Files），则安装成功后，需要在安装路径下找到 `dtempower.exe` 和 `dtempower_window.exe` 两个文件，并各自修改其文件属性，如图 43 所示。

如安装目录不需要管理员权限，用户可直接跳过此步操作。

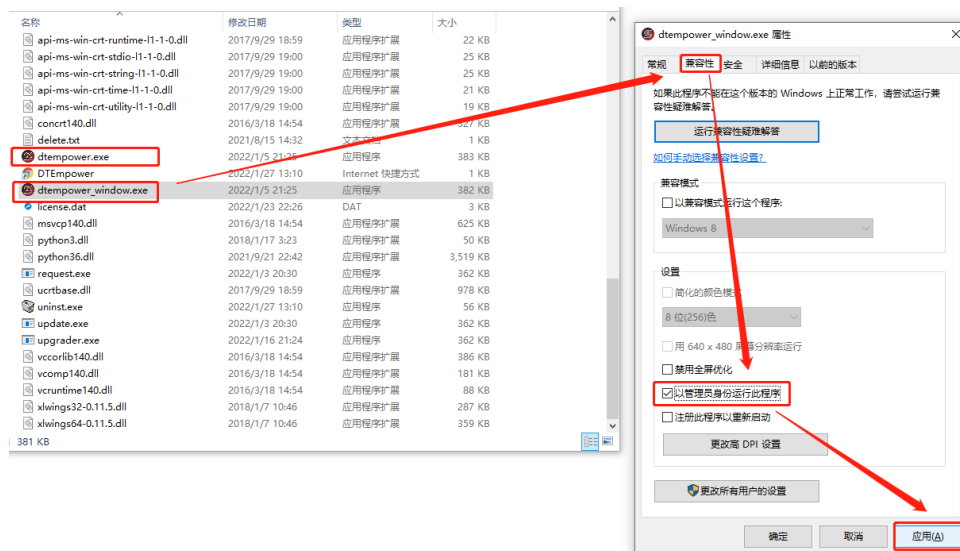


图43 以管理员身份运行程序



4.1.5 许可请求

找到 DTEmpower 的安装路径, 执行其中的 request.exe 程序(需要注意的是, 若安装路径在系统盘, 需要以管理员身份运行), 弹出用户信息配置弹窗, 如图 44、图 45 所示。

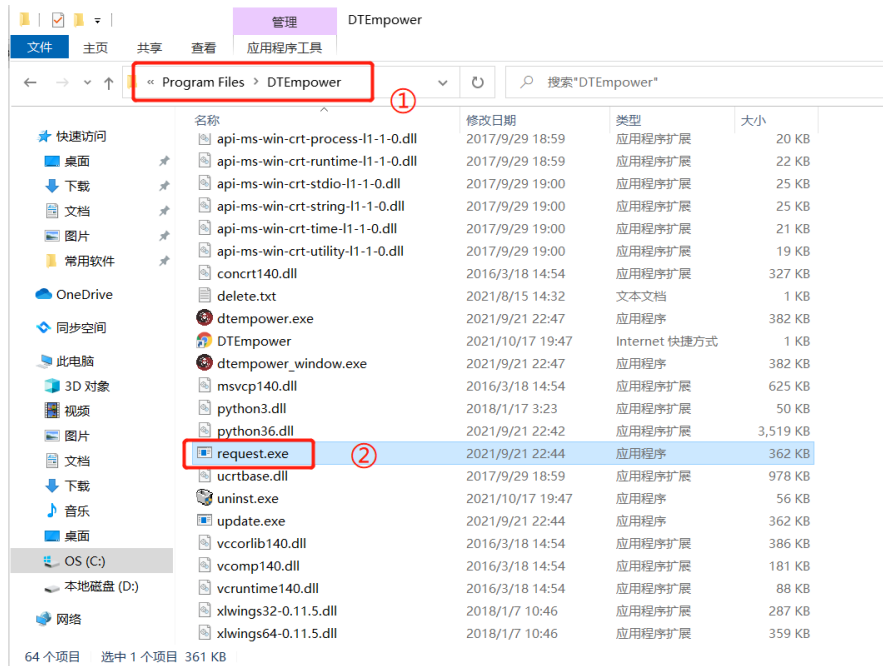


图44 生成许可请求文件

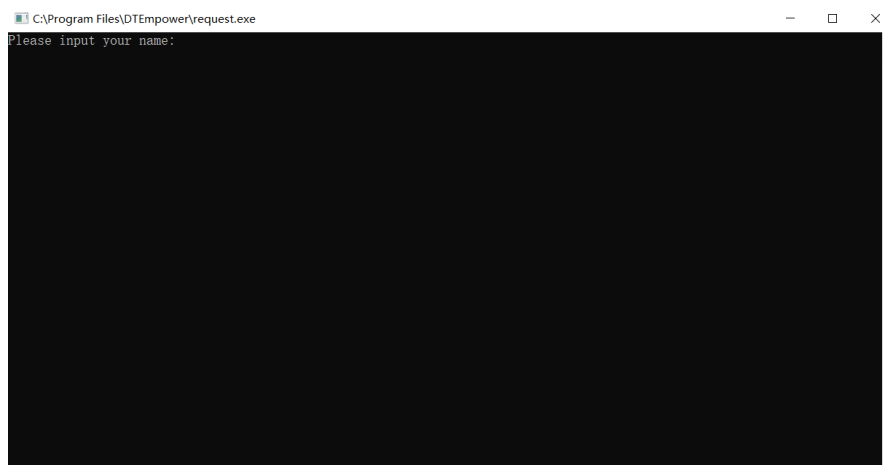


图45 许可请求弹窗

在窗口中录入用户名后单击【回车】键, 在安装路径下会生成与当前计算机硬件信息匹配的许可请求文件 request.txt 如图 46 所示。



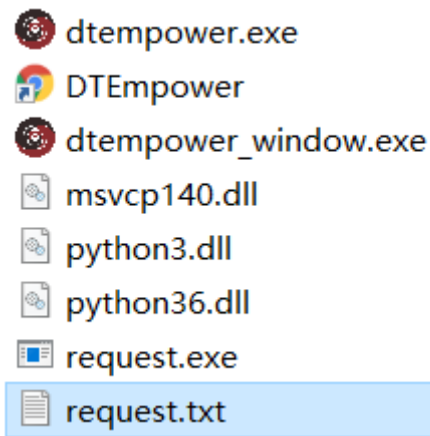


图46 生成许可请求文件

4.1.6 许可配置

将生成的许可请求文件发送给天沓软件，将传回的 license.dat 文件放置在软件安装路径根目录下如图 47 所示，即完成了许可文件的配置。

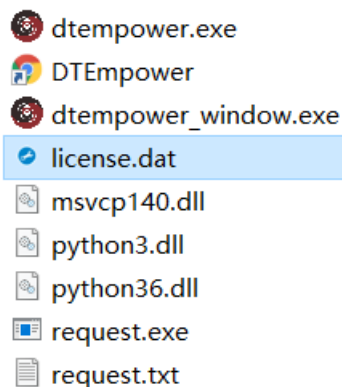


图47 授权 license

需要注意的是许可文件与生成许可请求文件时的计算机硬件匹配，若完成许可文件配置后存在更换主板、更换硬盘等情况，需要重新申请并配置许可文件。此外，许可文件使用时会记录每次的使用时间，服务器迁移等可能引发系统时间变更的操作会导致许可文件失效，需要重新申请并配置许可文件。

4.2 软件启动

DTEmpower 提供客户端启动和浏览器启动两种模式。其中客户端启动模式主要面向 DTEmpower 安装在本地计算机的情况；浏览器启动模式主要面向 DTEmpower 部署在服务器端或需要多学科人员协同合作的情况。



4.2.1 客户端启动

当 DTEmpower 安装在本地计算机时，可以采用客户端启动模式。客户端启动后，用户即可通过客户端窗口完成相关软件操作。



图48 客户端启动

DTEmpower 安装完成后，双击桌面上的 DTEmpower 快捷方式或安装目录下的 dtempower_window.exe 均可启动本地客户端，客户端启动成功后即进入项目新建/打开窗口，如图 48 所示，需要注意的是，若软件安装路径为系统盘等当前用户不具备读写权限的文件夹，启动时需要以管理员身份启动。

4.2.2 本地浏览器接入

用户在本地计算机打开浏览器（浏览器支持版本详见附录），在浏览器地址栏中录入“服务器端 IP:项目监听端口”后单击【回车】按钮即可实现浏览器接入，如图 49 所示。



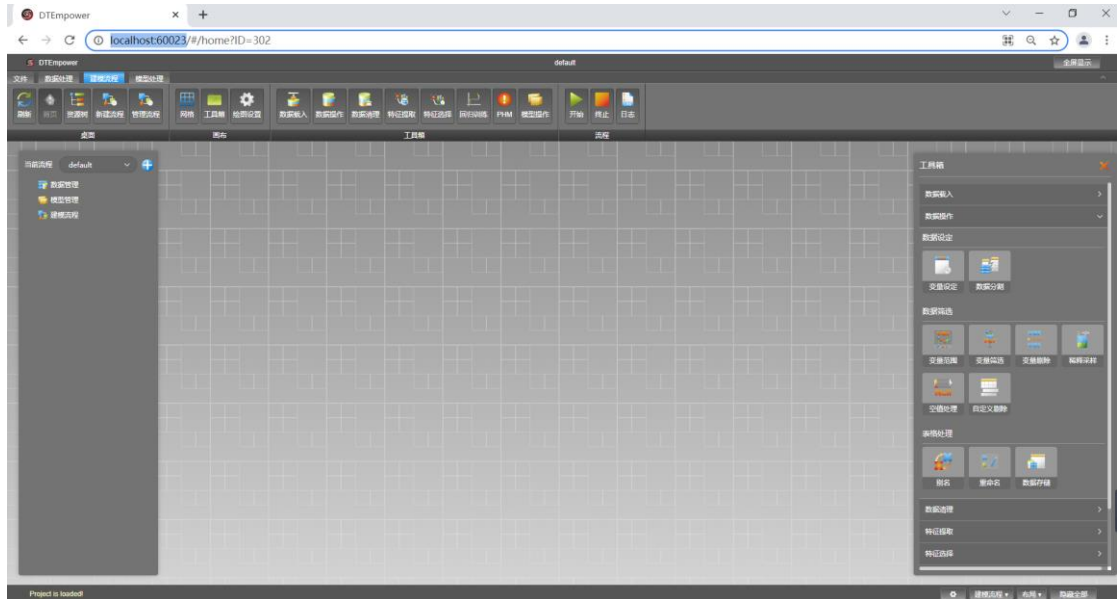


图49 本地浏览器接入

项目监听端口即图 49 命令窗口中的冒号后方的数字。需要注意的是，若同时打开多个项目，项目之间的端口不可相同。

4.3 软件欢迎页

在客户端模式下，软件启动后会首先进入欢迎页，如图 50 所示。欢迎页为用户提供了便利的指引功能，如快速打开或创建项目、浏览用户手册或示例项目等。

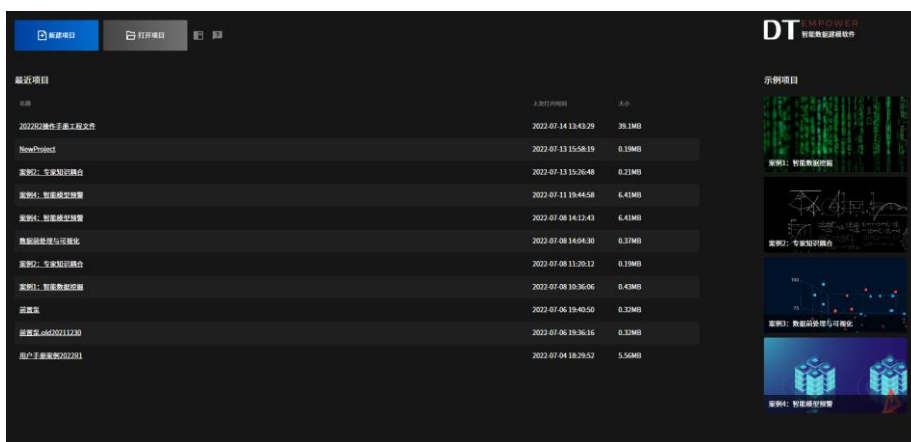


图50 软件欢迎页



4.3.1 新建项目

点击新建项目后，软件将自动关闭欢迎页并创建新的临时项目。新建的项目会开启独立的运行窗口，并显示空白画布，如图 51 所示。

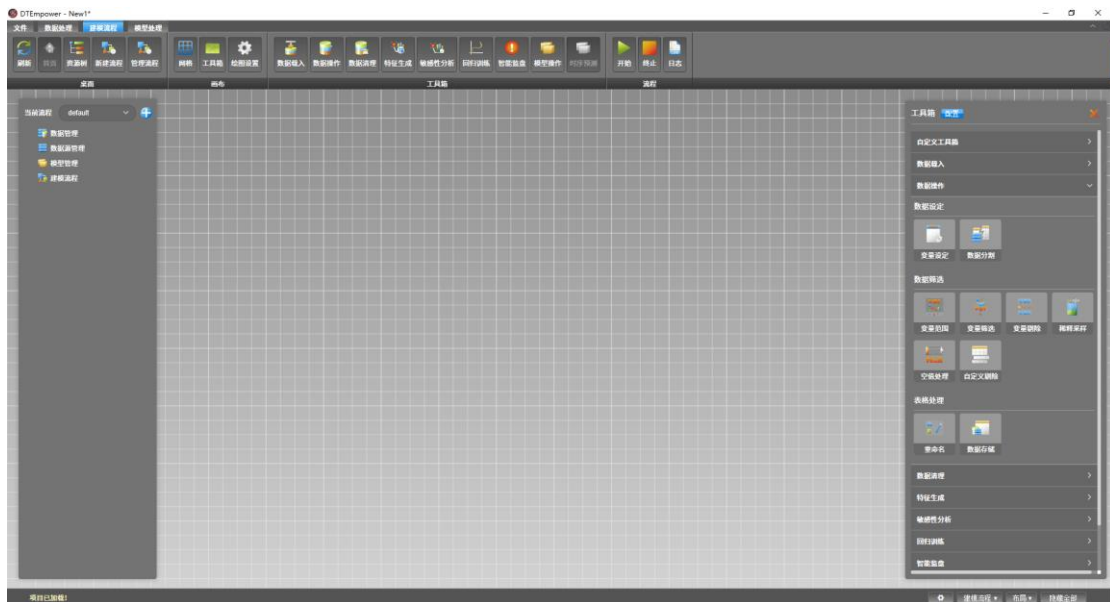


图51 新建项目

4.3.2 打开项目

点击打开项目按钮会弹出资源管理器弹窗，进入要打开的项目文件夹，选择其中的“.dtproj”（DTEmpower 存储项目文件格式），点击打开即可，如图 52 所示。

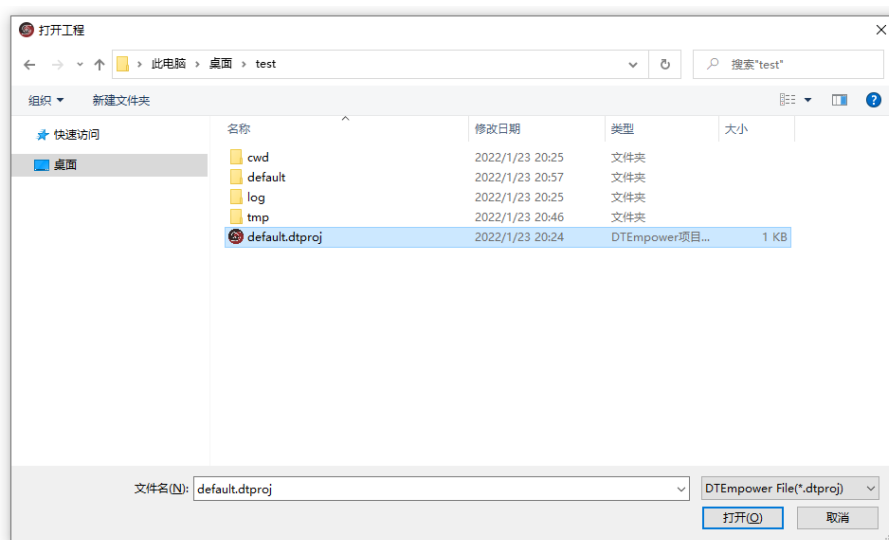


图52 打开项目

4.3.3 最近项目

在“最近项目”下，可查看近期所使用项目的名称、上次打开时间、以及项目文件大小。点击项目名称即可快速打开对应项目。

名称	上次打开时间	大小
2022R2操作手册工程文件	2022-07-11 19:20:55	2.55MB
案例4：智能模型预警	2022-07-08 14:12:43	6.41MB
数据前处理与可视化	2022-07-08 14:04:30	0.37MB
案例2：专家知识耦合	2022-07-08 11:20:12	0.19MB
案例1：智能数据挖掘	2022-07-08 10:36:06	0.43MB
前置菜	2022-07-06 19:40:50	0.32MB
前置菜_old20211230	2022-07-06 19:36:16	0.32MB
用户手册案例2022R1	2022-07-04 18:29:52	5.56MB

图53 最近项目

4.3.4 示例项目

在“示例项目”下，可查看软件内置的示例项目，点击对应卡片即可打开示例。新用户可基于示例项目体验界面交互、流程运行、结果查看等功能，实现快速上手。

注意：当软件安装在 C 盘时，需要使用管理员权限打开程序，之后才能正常运行示例项目功能。



图54 示例项目

4.3.5 使用帮助

点击图 55 中所示的按钮，即可自动打开用户操作手册（即本文档）。在操作手册中，详细描述了软件的功能、性能和用户界面。

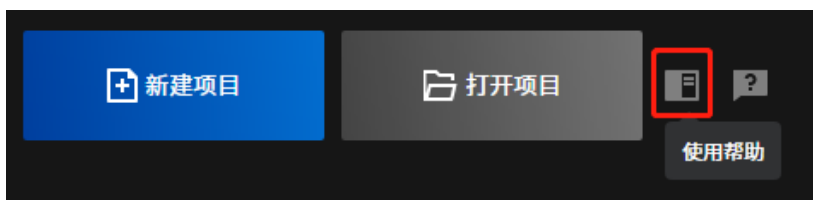


图55 使用帮助

4.3.6 问题反馈

点击图 56 中所示的按钮，即可使用默认浏览器打开 DTEmpower 官方网址。网站中提供新闻动态、原创文章、短视频等更多内容信息，以及天沓的联系方式，用户可通过此渠道反馈问题。

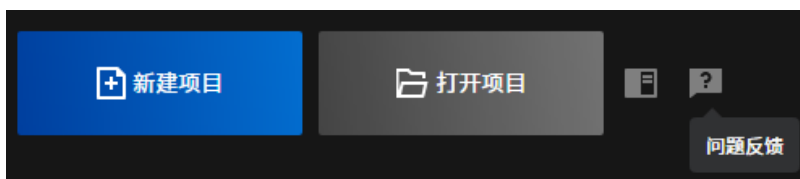


图56 问题反馈

4.4 项目管理

项目管理包含新建、保存与打开等操作。项目管理的进入方式为顶部 Ribbon 菜单栏的文件 Tab 页下的“项目”栏，如图 57 所示。



图57 项目管理

4.4.1 新建项目

点击新建项目后，软件将自动创建新的临时项目，如图 58 所示。新建的项目会开启独立的运行窗口，并显示空白画布。



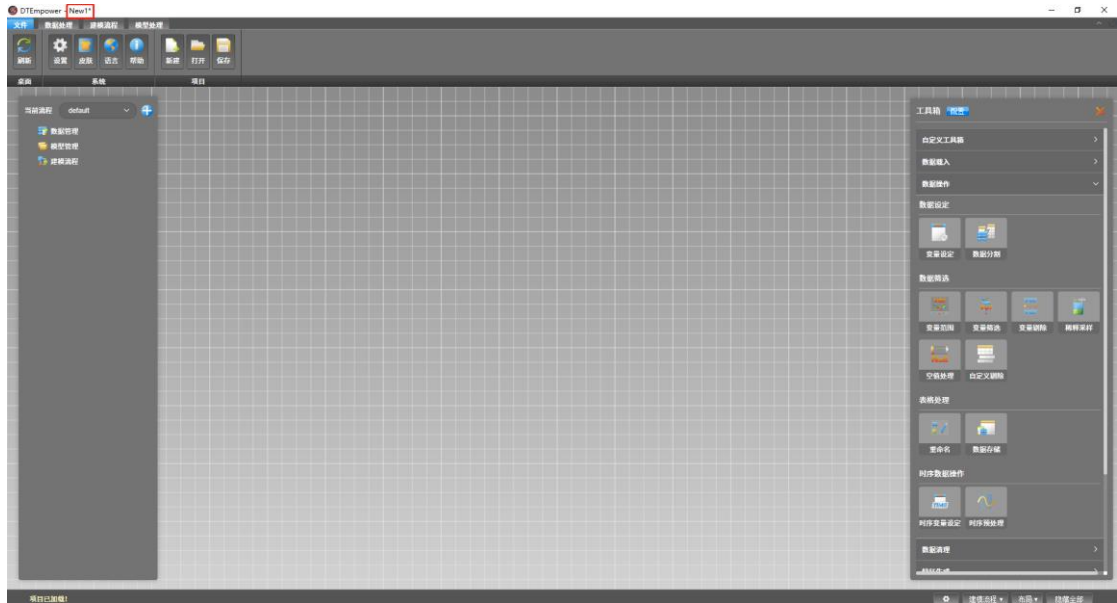


图58 新建项目

4.4.2 保存项目

新建的项目默认均为临时项目，当点击关闭后会被自动删除。通过保存项目操作可以将项目文件保存至本地，后续可以复用该项目。

通过点击菜单栏中的保存按钮，软件将跳出弹窗提示用户填写保存位置与工程名称，如图 59 所示。其中保存位置可以通过右侧“浏览”按钮选择本地的文件夹，下方的工程名称则用于填写保存的项目文件夹名称。



图59 保存项目

提示：在已经保存至本地的项目中，再次点击保存按钮会将项目文件转移至新的地址中。



4.4.3 打开项目

点击打开项目按钮会弹出资源管理器弹窗，进入要打开的项目文件夹，选择其中的“.dtproj”（DTEmpower 存储项目文件格式），点击打开即可，如图 60 所示。也可以在本地文件资源管理器中直接找到对应的项目文件，双击直接进入。

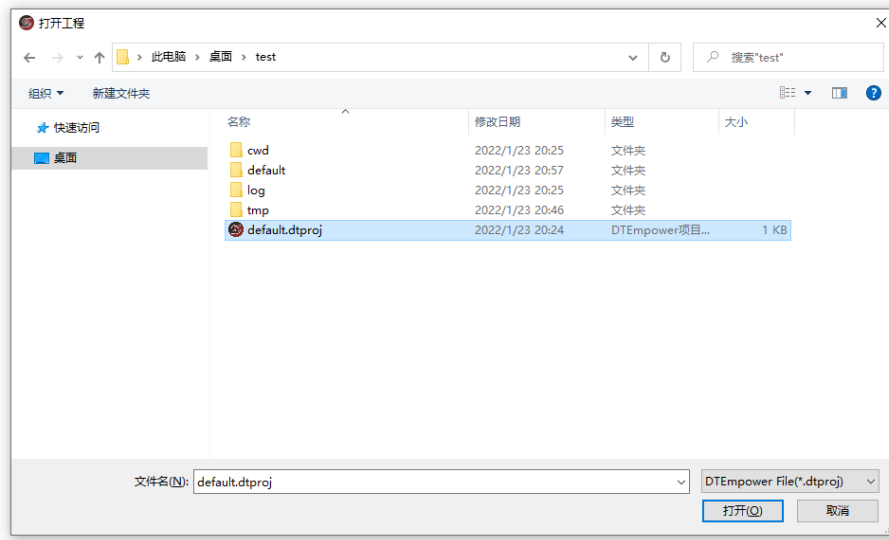


图60 打开项目

4.5 数据管理

4.5.1 上传数据集

在左侧资源树“数据管理”项上单击鼠标右键，即可弹出右键快捷菜单如图 61 所示。单击其中的上传按钮，即可弹出上传数据集的配置弹窗，如图 62。

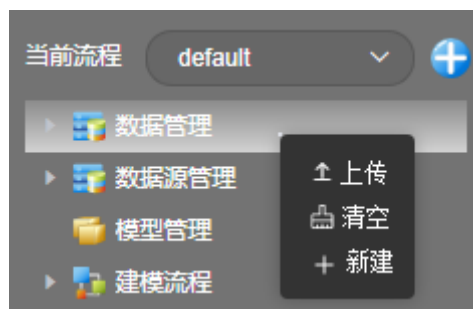


图61 上传数据集





图62 上传数据集配置弹窗

用户在弹窗中选择本地磁盘中的.csv、.xlsx、.xls、.dttable、南京天沅智能优化平台 AIPOD 的工程结果数据库文件后，单击确定按钮，即可完成数据集的上传，且自动匹配数据集名称与数据集类型，如图 63 所示。

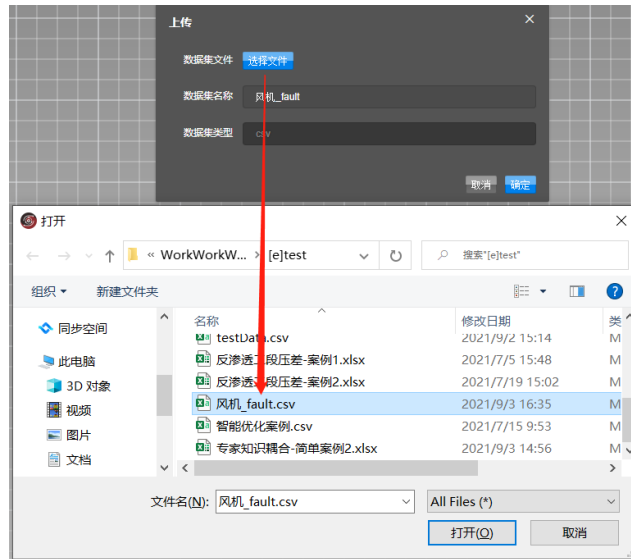


图63 选择本地磁盘数据文件上传

当数据集较大时，页面顶部会跳出软提示“数据集上传中，请稍等”，如图 64 所示。当数据集完成上传后，相同位置会出现绿色提示字样“数据上传成功”。

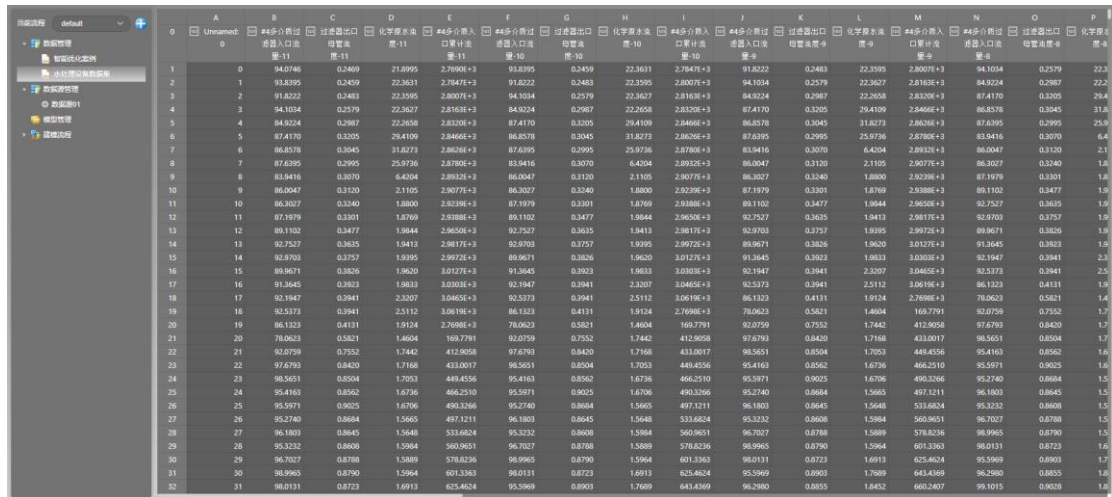


图64 大数据集上传等待提示



数据集上传完成后，以数据管理的子菜单形式存在，用户可单击数据管理菜单下的任意已添加的数据集，查看其数据信息，如图 65 所示。当数据集行列较多时，可以通过拖动底部和右侧的滑块进行全局查看。

注意：当数据文件格式为 csv、且单元格带有中文字符时，需要将文件编码转换为 UTF-8 格式，之后才能正常导入数据。



0	Unnamed	#4多小分过	过读进出口	D	E	F	G	H	I	J	K	L	M	N	O	P
		量-11	量-11	量-11	量-11	量-10	量-10	量-10	量-10	量-10	量-10	量-10	量-10	量-10	量-10	量-10
1	0	94.0746	0.2469	21.8995	2.7898E+3	93.8395	0.2459	22.3631	2.7847E+3	91.8222	0.2483	22.3595	2.8007E+3	94.1034	0.2579	22.3
2	1	93.8395	0.2459	22.3631	2.7847E+3	91.8222	0.2483	22.3595	2.8007E+3	94.1034	0.2579	22.3627	2.8163E+3	84.9224	0.2587	22.2
3	2	91.8222	0.2483	22.3595	2.8007E+3	94.1034	0.2579	22.3627	2.8163E+3	84.9224	0.2587	22.3628	2.8328E+3	87.4170	0.3205	29.4
4	3	94.1034	0.2579	22.3627	2.8163E+3	84.9224	0.2587	22.3628	2.8328E+3	87.4170	0.3205	29.4199	2.8466E+3	86.8578	0.3045	31.8
5	4	84.9224	0.2587	22.3628	2.8328E+3	87.4170	0.3205	29.4199	2.8466E+3	86.8578	0.3045	31.8273	2.8626E+3	87.6395	0.2995	25.0
6	5	87.4170	0.3205	29.4199	2.8466E+3	86.8578	0.3045	31.8273	2.8626E+3	87.6395	0.2995	25.9776	2.8781E+3	83.8416	0.3270	6.4
7	6	86.8578	0.3045	31.8273	2.8626E+3	87.6395	0.2995	25.9776	2.8781E+3	83.8416	0.3270	6.4204	2.8931E+3	86.0047	0.3130	2.1
8	7	87.6395	0.2995	25.9776	2.8781E+3	83.8416	0.3270	6.4204	2.8931E+3	86.0047	0.3130	2.1195	2.9077E+3	86.3027	0.3240	1.8
9	8	83.8416	0.3270	6.4204	2.8931E+3	86.0047	0.3130	2.1195	2.9077E+3	86.3027	0.3240	1.8800	2.9228E+3	87.1979	0.3301	1.8
10	9	86.0047	0.3130	2.1195	2.9077E+3	86.3027	0.3240	1.8800	2.9228E+3	87.1979	0.3301	1.8769	2.9388E+3	89.1102	0.3477	1.9
11	10	86.3027	0.3240	1.8800	2.9228E+3	87.1979	0.3301	1.8769	2.9388E+3	89.1102	0.3477	1.9844	2.9505E+3	92.7527	0.3635	1.9
12	11	87.1979	0.3301	1.8769	2.9388E+3	89.1102	0.3477	1.9844	2.9505E+3	92.7527	0.3635	1.9813	2.9617E+3	92.9703	0.3757	1.9
13	12	89.1102	0.3477	1.9844	2.9505E+3	92.7527	0.3635	1.9813	2.9617E+3	92.9703	0.3757	1.9925	2.9725E+3	93.8071	0.3826	1.9
14	13	92.7527	0.3635	1.9813	2.9617E+3	92.9703	0.3757	1.9925	2.9725E+3	93.8071	0.3826	1.9620	2.9827E+3	91.3645	0.3923	1.9
15	14	92.9703	0.3757	1.9925	2.9725E+3	93.8071	0.3826	1.9620	2.9827E+3	91.3645	0.3923	1.9833	2.9928E+3	92.1947	0.3941	2.3
16	15	89.9671	0.3826	1.9620	2.9827E+3	91.3645	0.3923	1.9833	2.9928E+3	92.1947	0.3941	2.3207	3.0035E+3	92.5773	0.3941	2.5
17	16	91.3645	0.3923	1.9833	2.9928E+3	92.1947	0.3941	2.3207	3.0035E+3	92.5773	0.3941	2.5112	3.0146E+3	86.1323	0.4131	1.9
18	17	92.1947	0.3941	2.3207	3.0035E+3	92.5773	0.3941	2.5112	3.0146E+3	86.1323	0.4131	1.9134	2.9996E+3	78.0623	0.5821	1.4
19	18	92.5773	0.3941	2.3112	3.0146E+3	86.1323	0.4131	1.9134	2.9996E+3	78.0623	0.5821	1.4664	1.60379E+3	92.9759	0.7252	1.3
20	19	86.1323	0.4131	1.9134	2.9996E+3	78.0623	0.5821	1.4664	1.60379E+3	92.9759	0.7252	1.7442	412.9058	87.6793	0.9430	1.2
21	20	78.0623	0.5821	1.4664	1.60379E+3	92.9759	0.7252	1.7442	412.9058	87.6793	0.9430	1.7448	433.0017	96.5651	0.6504	1.7
22	21	92.9759	0.7252	1.7442	412.9058	87.6793	0.9430	1.7448	433.0017	96.5651	0.6504	1.7053	449.4556	95.4163	0.8562	1.6
23	22	97.6793	0.8420	1.7168	433.0017	96.5651	0.6504	1.7053	449.4556	95.4163	0.8562	1.6736	466.2510	95.9771	0.9625	1.6
24	23	96.5651	0.6504	1.7053	449.4556	95.4163	0.8562	1.6736	466.2510	95.9771	0.9625	1.6796	498.3366	95.2740	0.8884	1.5
25	24	95.4163	0.8562	1.6736	466.2510	95.9771	0.9625	1.6796	498.3366	95.2740	0.8884	1.5685	497.1211	96.1803	0.8645	1.5
26	25	95.9771	0.9625	1.6796	498.3366	95.2740	0.8884	1.5685	497.1211	96.1803	0.8645	1.5646	533.6824	95.3332	0.8688	1.5
27	26	95.2740	0.8884	1.5685	497.1211	96.1803	0.8645	1.5646	533.6824	95.3332	0.8688	1.5984	565.9651	96.7027	0.8768	1.5
28	27	96.1803	0.8645	1.5646	533.6824	95.3332	0.8688	1.5984	565.9651	96.7027	0.8768	1.5889	578.8236	96.9965	0.8790	1.5
29	28	95.3332	0.8688	1.5889	565.9651	96.7027	0.8768	1.5889	578.8236	96.9965	0.8790	1.5964	601.3363	98.0131	0.8723	1.6
30	29	96.7027	0.8768	1.5889	578.8236	96.9965	0.8790	1.5964	601.3363	98.0131	0.8723	1.6913	625.4624	95.5969	0.8903	1.7
31	30	96.9965	0.8790	1.5964	601.3363	98.0131	0.8723	1.6913	625.4624	95.5969	0.8903	1.7069	643.4369	96.2983	0.8855	1.6
32	31	98.0131	0.8723	1.6913	625.4624	95.5969	0.8903	1.7069	643.4369	96.2983	0.8855	1.8452	662.2407	99.1015	0.9028	1.8

图65 表格查看数据信息

4.5.2 新建数据集

除了通过上传数据文件或连接数据库等方式建立数据集之外，用户还可以通过新建空白数据集的方式，从零开始搭建数据集。点击图 61 中的“新建”按钮，即可生成如图 66 所示的一行一列的空白数据集。之后用户可以通过添加更多的行与列，并将数据填充至各个单元格，最终获得所需要的数据集。

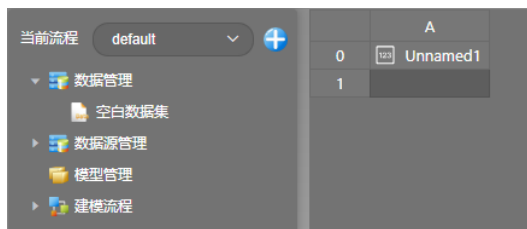


图66 新建空白数据集

4.5.3 复制数据集

当用户需要拷贝一份已有数据集的副本时，可以通过在资源树的相应数据集位置上点击鼠标右键，并选择其中的“复制”选项，如图 67 所示。由于数据分



析与数据建模过程中经常会涉及到数据集的反复调整与尝试，因此推荐事先拷贝多份数据集，并分别完成不同操作。

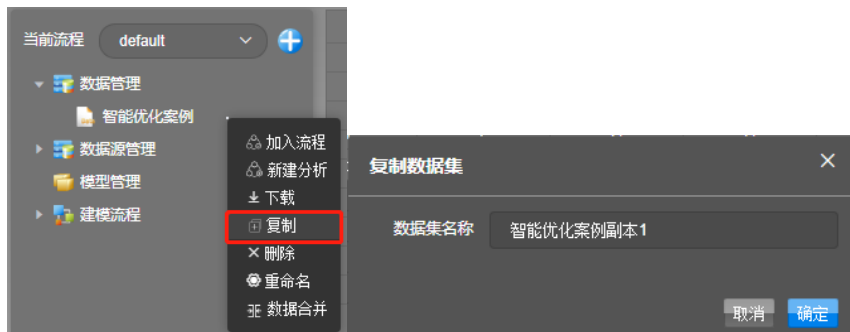


图67 复制数据集

4.5.4 下载数据集

提供两种下载数据集的入口，分别为在资源树-数据管理菜单下的任意数据集上单击鼠标右键、或在顶部菜单栏中点击下载，即可将对应数据集下载至本地磁盘中，如图 68 所示。

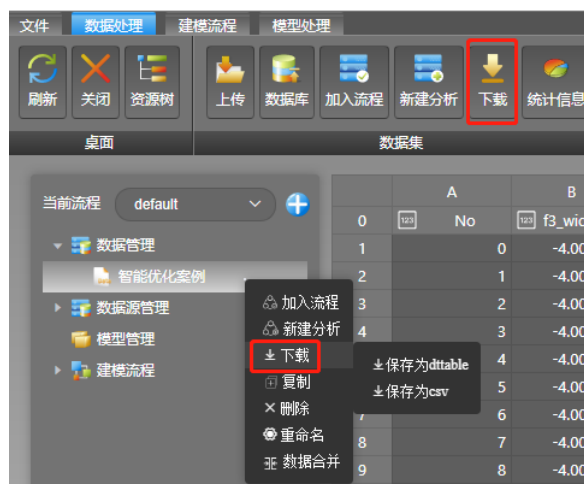


图68 下载数据集

数据集的下载选项提供了不同的保存形式。在 2022R2 版本中，支持将数据保存为 csv 文件（通用的文本格式文件）或 dttable 文件（DTEmpower 支持的数据集文件）。

4.5.5 重命名数据集

在资源树数据管理菜单下的任意数据集上单击鼠标右键，在右键快捷菜单中选择重命名按钮，即可修改对应数据集的名称，如图 69 所示。



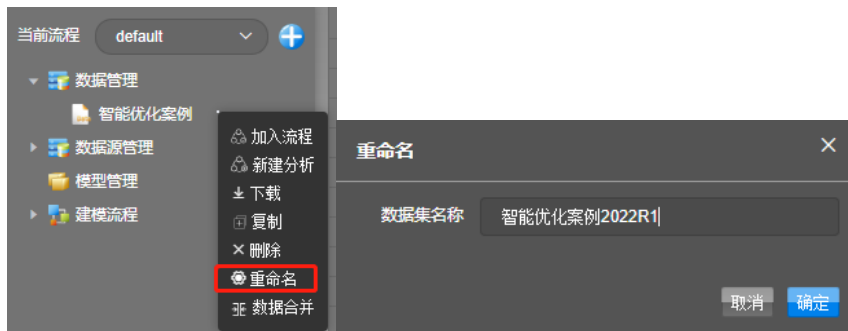


图69 重命名数据集

4.5.6 删除数据集

在资源树数据管理菜单下的任意数据集上单击鼠标右键，在右键快捷菜单中选择删除按钮，即可将对应数据集移除，如图 70 所示。注意此操作将永久删除该数据。

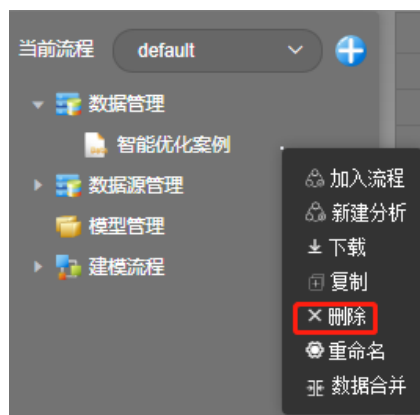


图70 删除数据集

4.5.7 清空数据管理

除逐个删除数据集外，在资源树数据管理的右键菜单中还提供有清空按钮，用以一键删除当前数据管理列表下的全部数据集，如图 71 所示。

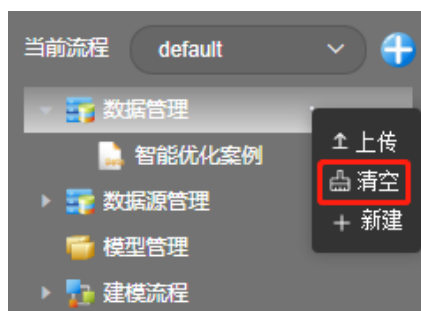


图71 清空数据管理



4.6 数据前处理

4.6.1 加载全部

在 DTEmpower 的数据管理模块中，上传的数据默认为只读模式，且仅显示前 500 行。如果需要开启数据前处理功能，其前提是点击“加载全部”按钮，一次性读入全部表格数据。“加载全部”按钮位于数据处理菜单下的前处理模块中，如图 72 所示。



图72 加载全部

当未开启“加载全部”时，数据集最大长度截断显示为 500 行，且仅支持部分表格操作功能（如复制），如图 73 所示。需要注意的是，此处截断显示只影响表格视图，而在实际建模流程、数据可视化、统计信息等环节中所使用的数据集仍为原始完整数据。

481	2019/1/7 9:17	19.5185	0.3341	978.0756	152.8604
482	2019/1/7 9:18	19.5625			52.8668
483	2019/1/7 9:19	19.5556			52.9245
484	2019/1/7 9:20	19.5255			53.0117
485	2019/1/7 9:21	19.5787			53.1008
486	2019/1/7 9:22	19.5602			53.1581
487	2019/1/7 9:23	19.5231			53.1992
488	2019/1/7 9:24	19.5440			53.2149
489	2019/1/7 9:25	19.5787			53.2471
490	2019/1/7 9:26	19.5926			53.2852
491	2019/1/7 9:27	19.6019			53.3172
492	2019/1/7 9:28	19.5509			53.3197
493	2019/1/7 9:29	19.5463			53.3382
494	2019/1/7 9:30	19.6019			53.3673
495	2019/1/7 9:31	19.5903			53.3798
496	2019/1/7 9:32	19.5509	0.3304	980.3087	153.3944
497	2019/1/7 9:33	19.5463	0.3304	980.4631	153.4392
498	2019/1/7 9:34	19.6019	0.3308	980.8519	153.4496
499	2019/1/7 9:35	19.5532	0.3306	981.2740	153.4838
500	2019/1/7 9:36	19.5718	0.3308	981.7155	153.4855



图73 未开启“加载全部”，默认截断显示前 500 行，且大部分前处理功能禁用
开启“加载全部”后，实际上实现两部分功能：

- a. 一次性读取并展示数据集中的全部数据；
- b. 开启数据前处理功能，如新增行列、设置数据格式、数据赋值等。

开启“加载全部”后的表格视图及右键菜单功能如图 74 所示。需要注意的是，由于“加载全部”功能会将全部数据一次性读入，对计算机的性能造成一定压力；因此对于大型数据集的修改，建议使用建模流程“数据操作”工具箱中的各类控件实现。

8841	2019/1/31 8:44	18.8470	0.3555	1.0314E+3
8842	2019/1/31 8:45	18.8380	0.3553	1.0313E+3
8843	2019/1/31	上方插入行	52	1.0309E+3
8844	2019/1/31	下方插入行	52	1.0309E+3
8845	2019/1/31	左方插入列	46	1.0302E+3
8846	2019/1/31	右方插入列	51	1.0303E+3
8847	2019/1/31	移除多行	56	1.0302E+3
8848	2019/1/31	移除该列	52	1.0301E+3
8849	2019/1/31	清空该列	59	1.0298E+3
8850	2019/1/31	剪切	52	1.0298E+3
8851	2019/1/31	复制	59	1.0297E+3
8852	2019/1/31	复制	55	1.0292E+3
8853	2019/1/31	变量格式	53	1.0293E+3
8854	2019/1/31	变量核查	52	1.0291E+3
8855	2019/1/31	数据赋值	59	1.0291E+3
8856	2019/1/31 8:59	18.8235	0.3567	1.0285E+3

图74 开启“加载全部”，可显示全部的数据，并支持丰富的前处理功能

4.6.2 通用表格操作

DTEmpower 的数据管理模块支持通用表格（行/列/单元格）操作。

4.6.2.1 选中

单选中：单击表格中的某个单元格，该单元格边框会变为蓝色，表明该单元格处于选中状态。此时该单元格可以接收来自键盘的输入以修改单元格内容（注意：修改的内容需点击“更新”按钮才会保存，下同）。

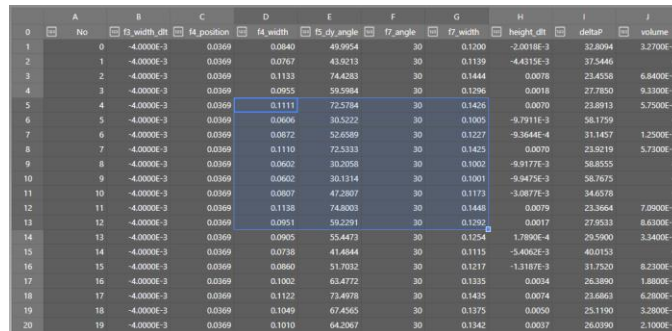
复数选中：按下键盘“Ctrl”键，连续单击表格中的多个单元格，即可实现复数个单元格的选中状态。



框选中：在某个单元格上按下鼠标左键不动，拖拽鼠标至另一个单元格，即可实现框选中其中所有单元格。同样的，单击表格中的某个单元格，按下键盘“Shift”键并单击另一个单元格，同样可以实现框选中其中所有单元格。

选中列：单击表格上方的列编号（A, B, C……），即可实现当前列内所有单元格的选中。与多选单元格类似，选中列支持“Ctrl”复选与“Shift”框选。

选中行：单击表格左侧的行编号（0, 1, 2, 3……），即可实现当前行内所有单元格的选中。与多选单元格类似，选中行支持“Ctrl”复选与“Shift”框选。注意：表头中的变量名称也作为特殊行存在，同样支持复制、粘贴等操作，并默认作为第 0 行显示。

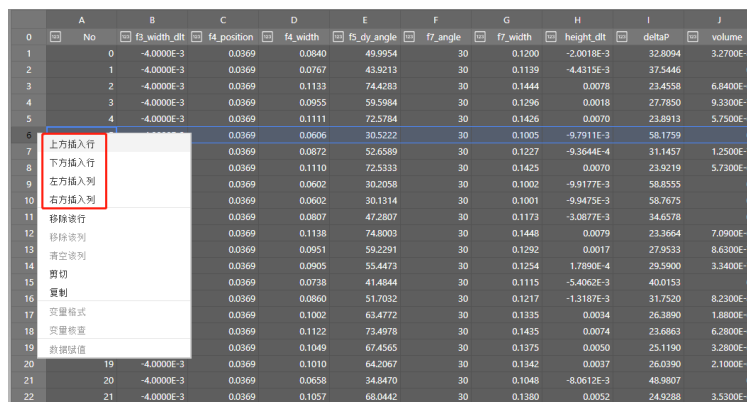


0	A	B	C	D	E	F	G	H	I	J
	No	f3_width_dlt	f4_position	f4_width	f5_dy_angle	f7_angle	f7_width	height_dlt	deltaP	volume
1	0	-4.000E-3	0.0369	0.0840	49.9954	30	0.1200	-2.0018E-3	32.8094	3.2700E-8
2	1	-4.000E-3	0.0369	0.0767	43.9213	30	0.1139	-4.4315E-3	37.5446	0
3	2	-4.000E-3	0.0369	0.1133	74.4283	30	0.1444	0.0078	23.4558	6.8400E-6
4	3	-4.000E-3	0.0369	0.0955	59.5984	30	0.1296	0.0018	27.7850	9.3300E-7
5	4	-4.000E-3	0.0369	0.1111	72.5784	30	0.1426	0.0070	23.8913	5.7500E-6
6	5	-4.000E-3	0.0369	0.0606	30.5222	30	0.1005	-9.7911E-3	58.1759	0
7	6	-4.000E-3	0.0369	0.0872	52.6589	30	0.1227	-9.3644E-4	31.1457	1.2500E-7
8	7	-4.000E-3	0.0369	0.1110	72.5333	30	0.1425	0.0070	23.9219	5.7300E-6
9	8	-4.000E-3	0.0369	0.0602	30.2058	30	0.1002	-9.9177E-3	58.8555	0
10	9	-4.000E-3	0.0369	0.0602	30.1314	30	0.1001	-9.9475E-3	58.7675	0
11	10	-4.000E-3	0.0369	0.0807	47.2807	30	0.1173	-3.0877E-3	34.6578	0
12	11	-4.000E-3	0.0369	0.1138	74.8003	30	0.1448	0.0079	23.3664	7.0900E-6
13	12	-4.000E-3	0.0369	0.0951	59.2291	30	0.1292	0.0017	27.9533	8.6300E-7
14	12	-4.000E-3	0.0369	0.0905	55.4472	30	0.1254	1.7890E-4	29.5900	3.3400E-7
15	14	-4.000E-3	0.0369	0.0738	41.4844	30	0.1115	-5.4062E-3	40.0153	0
16	15	-4.000E-3	0.0369	0.0860	51.7032	30	0.1217	-1.3187E-3	31.7520	8.2300E-8
17	16	-4.000E-3	0.0369	0.1002	63.4772	30	0.1335	0.0034	26.3890	1.8800E-6
18	17	-4.000E-3	0.0369	0.1122	73.4978	30	0.1435	0.0074	23.6863	6.2800E-6
19	18	-4.000E-3	0.0369	0.1049	67.4565	30	0.1375	0.0050	25.1190	3.2800E-6
20	19	-4.000E-3	0.0369	0.1010	64.2067	30	0.1342	0.0037	26.0390	2.1000E-6

图75 表格单元格选中状态

4.6.2.2 新增行列

在选中行（或列）的状态下，点击鼠标右键可见插入行（或列）的操作选项，如图 76 所示。新增的行列中所有单元格默认为空。特别的，对于新增列，会默认赋予列名“Unnamed”且指定为常规数值型变量格式（详见后文）。



0	A	B	C	D	E	F	G	H	I	J
	No	f3_width_dlt	f4_position	f4_width	f5_dy_angle	f7_angle	f7_width	height_dlt	deltaP	volume
1	0	-4.000E-3	0.0369	0.0840	49.9954	30	0.1200	-2.0018E-3	32.8094	3.2700E-8
2	1	-4.000E-3	0.0369	0.0767	43.9213	30	0.1139	-4.4315E-3	37.5446	0
3	2	-4.000E-3	0.0369	0.1133	74.4283	30	0.1444	0.0078	23.4558	6.8400E-6
4	3	-4.000E-3	0.0369	0.0955	59.5984	30	0.1296	0.0018	27.7850	9.3300E-7
5	4	-4.000E-3	0.0369	0.1111	72.5784	30	0.1426	0.0070	23.8913	5.7500E-6
6	5	-4.000E-3	0.0369	0.0606	30.5222	30	0.1005	-9.7911E-3	58.1759	0
7	6	-4.000E-3	0.0369	0.0872	52.6589	30	0.1227	-9.3644E-4	31.1457	1.2500E-7
8	7	-4.000E-3	0.0369	0.1110	72.5333	30	0.1425	0.0070	23.9219	5.7300E-6
9	8	-4.000E-3	0.0369	0.0602	30.2058	30	0.1002	-9.9177E-3	58.8555	0
10	9	-4.000E-3	0.0369	0.0602	30.1314	30	0.1001	-9.9475E-3	58.7675	0
11	10	-4.000E-3	0.0369	0.0807	47.2807	30	0.1173	-3.0877E-3	34.6578	0
12	11	-4.000E-3	0.0369	0.1138	74.8003	30	0.1448	0.0079	23.3664	7.0900E-6
13	12	-4.000E-3	0.0369	0.0951	59.2291	30	0.1292	0.0017	27.9533	8.6300E-7
14	12	-4.000E-3	0.0369	0.0905	55.4472	30	0.1254	1.7890E-4	29.5900	3.3400E-7
15	14	-4.000E-3	0.0369	0.0738	41.4844	30	0.1115	-5.4062E-3	40.0153	0
16	15	-4.000E-3	0.0369	0.0860	51.7032	30	0.1217	-1.3187E-3	31.7520	8.2300E-8
17	16	-4.000E-3	0.0369	0.1002	63.4772	30	0.1335	0.0034	26.3890	1.8800E-6
18	17	-4.000E-3	0.0369	0.1122	73.4978	30	0.1435	0.0074	23.6863	6.2800E-6
19	18	-4.000E-3	0.0369	0.1049	67.4565	30	0.1375	0.0050	25.1190	3.2800E-6
20	19	-4.000E-3	0.0369	0.1010	64.2067	30	0.1342	0.0037	26.0390	2.1000E-6
21	20	-4.000E-3	0.0369	0.0858	34.8470	30	0.1048	-8.0612E-3	48.9807	0
22	21	-4.000E-3	0.0369	0.1057	68.0442	30	0.1380	0.0052	24.9288	3.5300E-6

图76 新增行列



4.6.2.3 移除行列

在选中行（或列）的状态下，点击鼠标右键可见移除该行（或列）的操作选项，如图 77 所示。同时选中复数个行（或列）时，同样支持移除操作。

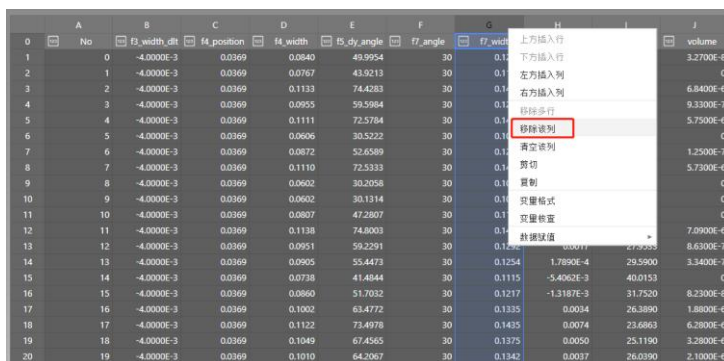


图77 移除行列

4.6.2.4 清空行列

在选中行（或列）的状态下，点击鼠标右键可见清空该行（或列）的操作选项，如图 78 所示。与移除行列不同，清空操作会保留选中行（或列）的位置，并将其中所有单元格置空。同时选中复数个行（或列）时，同样支持清空操作。

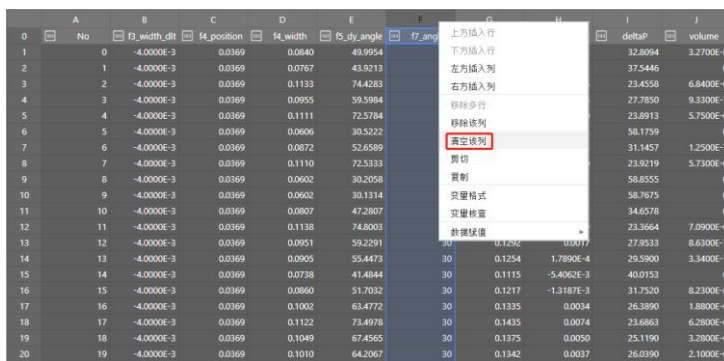


图78 清空行列

4.6.2.5 剪切、复制与粘贴

在选中状态（单元格/行/列）下，点击鼠标右键可见剪切与复制的操作选项，如图 79 所示。由于浏览器的安全限制，粘贴按钮未以显示按钮的形式加入右键菜单中，但支持键盘“Ctrl+V”的快捷键。类似的，剪切与复制操作同样支持“Ctrl+X/C”的快捷键。DTEmpower 表格的复制和粘贴功能支持与 MS Excel 等其他软件中的表格数据互通。



0	A	B	C	D	E	F	G	H	I	J
No	f3_width_dlt	f4_position	f4_width	f5_dy_angle	f7_angle	f7_width	height_dlt	deltap	volume	
1	0	-4.000E-3	0.0369	0.0840	49.9954	30	0.1200	-2.0018E-3	32.8094	3.2700E-8
2	1	-4.000E-3	0.0369	0.0767	43.9213	30	0.1139	-4.4315E-3	37.5446	0
3	2	-4.000E-3	0.0369	0.1133	74.4283	30	0.1444	0.0078	23.4558	6.8400E-6
4	3	-4.000E-3	0.0369	0.0955	59.5984	30	0.1296	0.0018	27.7850	9.3300E-7
5	4	-4.000E-3	0.0369	上方插入行		30	0.1426	0.0070	23.8913	5.7500E-6
6	5	-4.000E-3	0.0369	下方插入行		30	0.1005	-9.7911E-3	58.1759	0
7	6	-4.000E-3	0.0369	左方插入列		30	0.1227	-9.3644E-4	31.1457	1.2500E-7
8	7	-4.000E-3	0.0369	右方插入列		30	0.1425	0.0070	23.9219	5.7300E-6
9	8	-4.000E-3	0.0369	删除该行		30	0.1002	-9.9177E-3	58.8555	0
10	9	-4.000E-3	0.0369	删除该列		30	0.1001	-9.9475E-3	58.7675	0
11	10	-4.000E-3	0.0369	固定该列		30	0.1173	-3.0877E-3	34.6578	0
12	11	-4.000E-3	0.0369	剪切		30	0.1448	0.0079	23.3664	7.0900E-6
13	12	-4.000E-3	0.0369	复制		30	0.1292	0.0017	27.9533	8.6300E-7
14	13	-4.000E-3	0.0369	复制		30	0.1254	1.7890E-4	29.5900	3.3400E-7
15	14	-4.000E-3	0.0369	变量格式		30	0.1115	-5.4062E-3	40.0153	0
16	15	-4.000E-3	0.0369	变量重置		30	0.1217	-1.3187E-3	31.7520	8.2300E-8
17	16	-4.000E-3	0.0369	删除该值		30	0.1335	0.0034	26.3890	1.8800E-6
18	17	-4.000E-3	0.0369			30	0.1435	0.0074	23.6863	6.2800E-6

图79 剪切、复制与粘贴

4.6.2.6 重命名变量

在数据管理模块的表格查看功能中，变量名称作为特殊行（第 0 行）存在，支持在单元格选中状态下直接接收键盘输入修改，同时还支持批量粘贴覆盖操作。

4.6.2.7 调整变量顺序

在数据管理模块的表格查看功能中，支持在列选中的状态下，按下鼠标左键并拖拽，即可向前后调整变量顺序。

4.6.2.8 调整列宽

内置表格查看功能固定了默认的列宽。当单元格内容较长时，可以在表头右边缘（如图所示）按下鼠标左键并前后拖动，即可调整列宽。此外，还支持在相应位置双击鼠标左键进行列宽自适应。

0	A	B	C	D	E	F	G	H	I	J
No	f3_width_dlt	f4_position	f4_width	f5_dy_angle	f7_angle	f7_width	height_dlt	f7_angle	deltap	volume
1	0	-4.000E-3	0.0369	0.0840	49.9954	0.1308	-2.0018E-3	30	32.8094	3.2700E-8
2	1	-4.000E-3	0.0369	0.0767	43.9213	0.1139	-4.4315E-3	30	37.5446	0
3	2	-4.000E-3	0.0369	0.1133	74.4283	0.1444	0.0078	30	23.4558	6.8400E-6
4	3	-4.000E-3	0.0369	0.0955	59.5984	0.1296	0.0018	30	27.7850	9.3300E-7
5	4	-4.000E-3	0.0369	0.1111	72.5784	0.1426	0.0070	30	23.8913	5.7500E-6
6	5	-4.000E-3	0.0369	0.0906	30.5222	0.1005	-9.7911E-3	30	58.1759	0
7	6	-4.000E-3	0.0369	0.0872	52.6589	0.1227	-9.3644E-4	30	31.1457	1.2500E-7
8	7	-4.000E-3	0.0369	0.1110	72.5333	0.1425	0.0070	30	23.9219	5.7300E-6
9	8	-4.000E-3	0.0369	0.0902	30.2058	0.1002	-9.9177E-3	30	58.8555	0
10	9	-4.000E-3	0.0369	0.0902	30.1314	0.1001	-9.9475E-3	30	58.7675	0
11	10	-4.000E-3	0.0369	0.0807	47.2807	0.1173	-3.0877E-3	30	34.6578	0
12	11	-4.000E-3	0.0369	0.1138	74.8003	0.1448	0.0079	30	23.3664	7.0900E-6
13	12	-4.000E-3	0.0369	0.0951	59.2291	0.1292	0.0017	30	27.9533	8.6300E-7
14	13	-4.000E-3	0.0369	0.0905	55.4473	0.1254	1.7890E-4	30	29.5900	3.3400E-7
15	14	-4.000E-3	0.0369	0.0738	41.4844	0.1115	-5.4062E-3	30	40.0153	0

图80 调整列宽

4.6.2.9 撤销

在数据管理模块的表格查看功能中，支持利用快捷键“Ctrl+Z”完成单步撤销操作。连续多次应用该快捷键可以实现多步撤销功能。类似的，快捷键“Ctrl+Y”支持对撤销操作的撤销，且同样支持多步操作。需要注意的是，并非所有的数据前处理操作均可以实现撤销功能，如列的新增与删除等改变表格结构的操作。



4.6.3 更新与撤回

在数据管理的前处理操作中，数据集存在两种状态，分别为“确认状态”与“待更新状态”。二者在资源树上的显示状态有所区别，分别呈现彩色图标与灰色图标，如图 81 所示。二者的区别在于，后者含有未保存确认的用户前处理操作，前者则已保存最新数据。

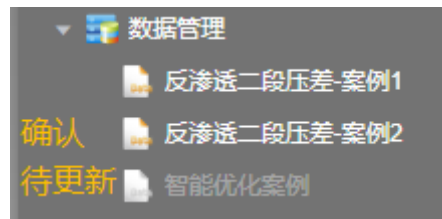


图81 数据集更新状态

确认状态与待更新状态之间可以相互转化。对于原本处于确认状态的数据集，用户做了任意前处理操作，如更改单元格数值、移动变量位置等，均会导致数据集变更为待更新状态。而待更新状态的数据集如需转变为确认状态，需要通过 Ribbon 菜单栏的“更新”或“撤回”按钮，如图 82 所示。其中，更新操作将保存用户所做的全部修改，撤回操作则将取消全部修改并返回上一个确认状态。对于处于确认状态的数据集，更新与撤回按钮不可用。

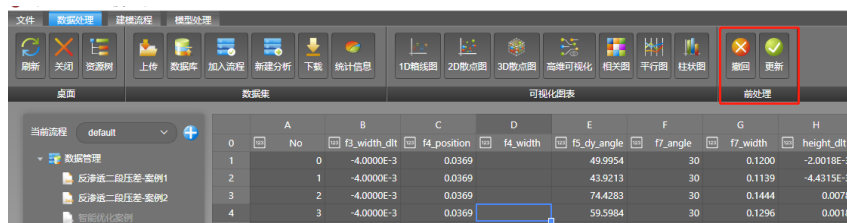


图82 更新与撤回按钮

注意：在处理如数据可视化、数据合并、加入流程等操作时，要求数据集处于确认状态。如存在未经确认的更改操作，将弹窗提示用户是否要保留更改操作。

4.6.4 变量格式

在 DTEmpower 的数据管理表格中，每个变量（即表格中的每一列）都需要明确隶属于某一种变量格式，且变量下的所有数据（即列下的所有单元格）格式应与变量格式保持一致。变量格式的设定非常重要，会影响后续的数据审查、数据可视化、建模流程等多个环节。



规定所有变量有 3 种基本格式：数值型、字符型和日期型。其中每种基本格式下又会进一步细分，共包含 10 种不同的变量格式，分别为常规数值、整型、浮点型、百分比、科学计数、有序分类、无序分类、纯文本、时间型和日期型。更改变量格式的入口如图 83 所示，在对应列的表头位置单击鼠标右键，即可看到对应选项。

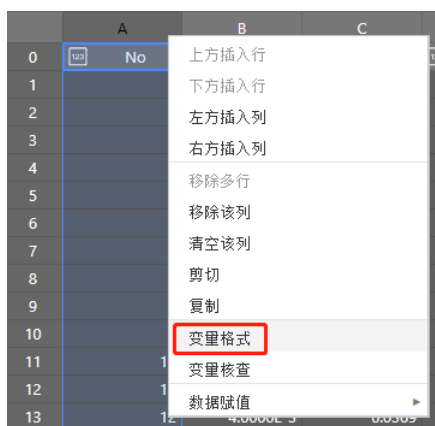


图83 变量格式

4.6.4.1 常规数值

常规数值型变量格式中不包含任何特定的数字格式，并用统一的方式进行显示。常规数值是默认的数值型变量格式。导入数据文件会优先识别为常规数值类型，且新增列的默认变量格式同样为常规数值型。



图84 常规数值型变量

常规数值型变量格式定义了一套统一的表格显示形式。对于整型数据将保持默认显示，而浮点型数据默认保留 4 位小数，且当数据位数过长时会修改为科学计数法显示。



4.6.4.2 整型

在整型变量格式中，单元格支持存储与处理整数类型的数据。



图85 整型变量

4.6.4.3 浮点型

在浮点型变量格式中，单元格支持存储与处理浮点数类型的数据。用户还可以指定显示的小数位数。提示：小数位数的设置只影响表格显示，不改变实际计算过程中的精度。



图86 浮点型变量

4.6.4.4 百分比

在百分比变量格式中，单元格内容以百分比的形式显示，即实际数值乘以 100。用户可以指定显示的小数位数。与浮点型格式一致，小数位数的设置只影响表格显示，不改变实际计算过程中的精度。





图87 百分比变量

4.6.4.5 科学计数

在科学计数变量格式中，单元格内容以科学计数法的形式显示，其中 E 表示 10 的幂。用户可以指定显示的小数位数。与浮点型格式一致，小数位数的设置只影响表格显示，不改变实际计算过程中的精度。



图88 科学计数变量

4.6.4.6 有序分类

有序分类是一种特殊的字符型变量格式。有序分类变量格式既包含类别的信息，还包含次序的信息，分类值支持排序但不支持算术运算。





图89 有序分类变量

在有序分类变量中，引入了分类标签和分类值 2 类概念，且二者一一对应。其中分类标签是在表格单元格中的外部显示，而分类值是每种分类标签的定量化描述。当修改变量格式为有序分类时，可以通过“自动匹配”按钮将当前列中所有不同的类型各单独分配一组分类标签和分类值。同时分类设定中还具有添加、删除、移动分类类型的功能，用以手动匹配。

注意：为保证分类有意义，DTEmpower 中限制不同分类类型的个数不能超过 20 种。

4.6.4.7 无序分类

无序分类变量格式对应数据的属性进行分类，分类值仅代表类别差异不表示大小关系。无序分类变量格式设定的其他操作方式与有序分类设定基本一致。



图90 无序分类变量



4.6.4.8 纯文本

纯文本变量格式将单元格内容作为文本处理，显示内容与输入内容完全一致。纯文本格式不限制单元格的输入内容，但相应的可以实现的功能也相对较弱。



图91 纯文本变量

4.6.4.9 时间型

时间型变量格式将单元格内数据识别为时间长度，时间单位由用户指定。时间型变量单元格中存储的数据本质仍为数值，但支持与日期型变量的加减运算。



图92 时间型变量

4.6.4.10 日期型变量

日期型变量格式将单元格内数据识别为用户指定的日期格式。一般的，新导入数据会自动将其中的日期列识别为日期型变量，且支持的日期格式如图中下拉弹窗所示。在范围外的特殊日期格式（如 2021-01/01），需用户手动调整。





图93 日期型变量

4.6.5 变量核查

变量核查的功能入口如图 94 所示，需在选中某数据列的状态下，在表头唤出右键菜单并点击其中的变量核查选项。

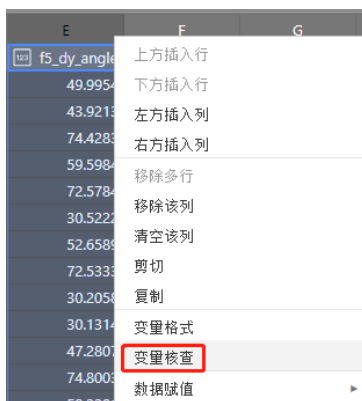


图94 变量核查

变量范围共分为前后 2 步操作：首先需确定需要核查的变量行列范围与核查规则，之后对核查出的异常值进行跳转查看、（批量）删除或（批量）替换等操作。需要注意的是，变量核查的功能与选定的变量格式有密切关联。

4.6.5.1 检查范围

在检查范围中需要用户指定要完成核查的表格行列范围。其中列范围规定每次仅能对单列（即单变量）完成检查。行范围默认选择首行与尾行，用户也可以根据需要限定局部范围。



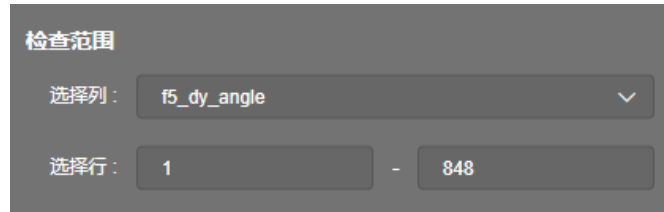


图95 变量核查范围

4.6.5.2 检查规则

根据所选定的变量格式，提供了不同的检查规则选项。以数值型变量格式为例，提供了 5 种不同的检查规则：空值、重复值、变量格式不匹配、区间范围与单调性，如图 96 所示。



图96 变量核查规则

- 空值：单元格中内容为空。
- 重复值：有两个及以上的单元格中内容相同。
- 变量格式不匹配：单元格中内容格式与表头所设定的变量格式不匹配。
- 区间范围：出现下级选项菜单，定义范围类型及输入框；可选的类型包括介于、不介于、小于、小于等于、大于、大于等于、等于和不等于等。
- 单调性：出现下级选项菜单，定义单调性类型；可选的类型包括严格单调递增、不严格单调递增、严格单调递减、不严格单调递减。

提示 1：空值、重复值、变量格式不匹配、区间范围等 4 种检查规则会直接筛选出符合用户所设定的检查规则，而单调性检查规则会筛选出不符合单调性规律的异常数值。用户在使用时需注意辨别其中的区别。



提示 2：变量格式会影响检查规则的类型选项和输入框方式等。如无序分类变量只包含空值、变量格式不匹配、区间范围三种检查规则。时间型变量的输入框会提供时间选择器等。

确认了检查范围与检查规则之后，点击确定即可筛选出对应的异常值列表。进一步的，提供了查找、删除、替换三个 Tab 页用于完成核查后处理。

4.6.5.3 核查处理-查找

查找处理页提供了逐个查看或跳转查看异常数据的功能，如图 97 所示。查找上一个、查找下一个会在异常值列表中循环，并在表格中给出对应选中状态。在跳转查看功能中，既显示了异常值的个数，用户还可直接输入序号跳转到对应位置。

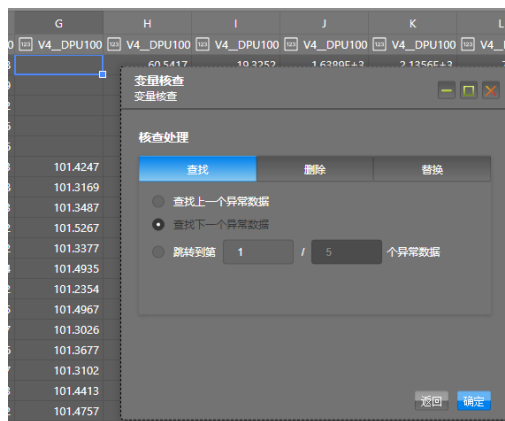


图97 核查处理-查找

4.6.5.4 核查处理-删除

删除处理页提供了逐个删除行与删除全部行的功能，如图 98 所示。用户可以通过窗口底部的“上一个”或“下一个”按钮进行选中异常值列表的切换。





图98 核查处理-删除

4.6.5.5 核查处理-替换

替换处理页在提供逐个异常值替换和全部异常值批量替换功能的基础上，还增加了替换方式的选项，包括列均值、前后插值和固定值等，如图 99 所示。



图99 核查处理-替换

- 列均值：将异常值所在单元格填充为当前列全部有效值的平均值；
- 前后插值：将异常值所在单元格填充为相邻有效值的加权平均值（取行序号作为权重），首尾单元格则与最近的有效值保持一致；
- 固定值：将异常值所在单元格填充为用户指定的数值；如为日期型变量提供时间选择器。

4.6.6 数据赋值

数据赋值的功能入口如图 100 所示，需在选中某数据列的状态下，在表头唤出右键菜单并点击其中的数据赋值选项。



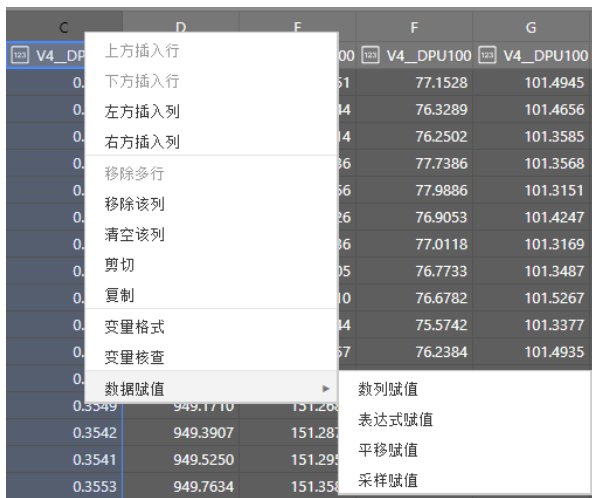


图100 数据赋值

DTEmpower 中共提供了 4 种数据赋值的方式：数列赋值、表达式赋值、平移赋值以及采样赋值。

4.6.6.1 数列赋值

当目标变量是依据一定规律生成的数列时，可以使用数列赋值的方式进行生成。数列赋值需要用户指定三方面内容：

- 目标变量：指定要完成赋值的变量列列名；
- 数列类型：指定目标变量满足的规律，共支持 5 种形式的数列类型，分别为常数列、等差数列、等比数列、随机数列以及周期数列，如图 101 所示；



图101 定义数列类型

- 初值：根据不同数列类型，需提供相应的初值。数列类型与初值的对应关系如表 1 所示。



表1 数列类型与初值对应表

数列类型	初值
常数数列	固定值：指定所有单元格存储相同固定值
等差数列	首项：指定等差数列的首项 公差：指定等差数列的公差
等比数列	首项：指定等比数列的首项 公差：指定等比数列的公比
随机数列	区间下限：指定生成随机数最小值 区间上限：指定生成随机数最大值 随机种子：指定随机数生成的随机种子
周期数列	周期赋值：指定循环周期中的全部数据

4.6.6.2 表达式赋值

当目标变量是由其他变量通过一定的函数变换生成时，可以使用表达式赋值的方式。表达式赋值的页面如图 102 所示。

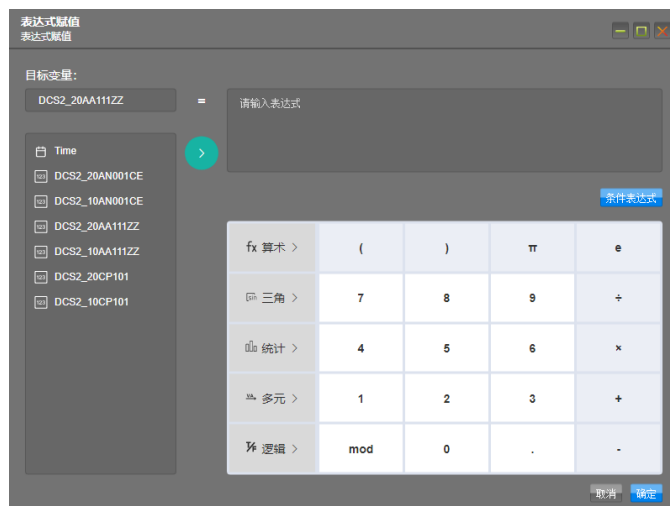


图102 表达式赋值

表达式赋值首先需指定目标变量。目标变量可以从已有变量中选择，也可以输入新列名并增加新列。

右上角的表达式输入框定义了目标变量的函数表达式，除直接接收键盘输入外，还可以与左侧的变量列表区、下方的计算器表盘进行联动。表达式赋值既支持简单的四则运算，还支持复杂的向量、函数与条件运算，具体如下：

- 四则运算：如 $y = 8 \times \pi + e$ ，当检测到计算结果为标量时，会为所有单元格赋值相同的数值；



- 向量运算：如 $y = x_1 + x_2$ ，当检测到计算结果为向量时，会为所有单元格按照向量顺序依次赋值；需要注意的是，向量之间的乘（除）默认为逐元素相乘（除）；用户可以从左侧的变量列表区选择列作为向量运算的输入；

- 函数运算：在下方的计算器表盘中支持多种函数运算，如算术函数、三角函数、统计函数、多元函数、逻辑函数等计算；

- 条件运算：即分段函数，由两部分组成：条件语句与表达式，二者一一对应。开启条件运算的条件是点击弹窗中高亮的“条件表达式”按钮，并进入新的弹窗，如图 103 所示。该弹窗的形式与使用方法与图 102 中的弹窗基本一致，但在右上角相同的位置上需要用户输入条件语句，其输出值为布尔值（True/False）；可以通过左侧的“+”按钮建立多个条件语句；当用户输入了条件语句后，点击“继续”按钮可以返回初始的表达式赋值窗口，并会在输入栏下方看到已经建立的分段条件语句（其他为自动生成），如图 104 所示。切换不同的条件语句会同步切换不同的表达式输入框。

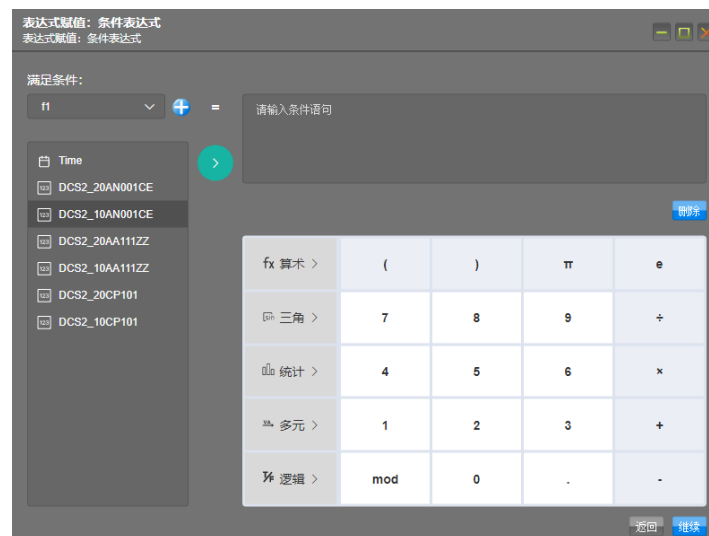


图103 条件表达式



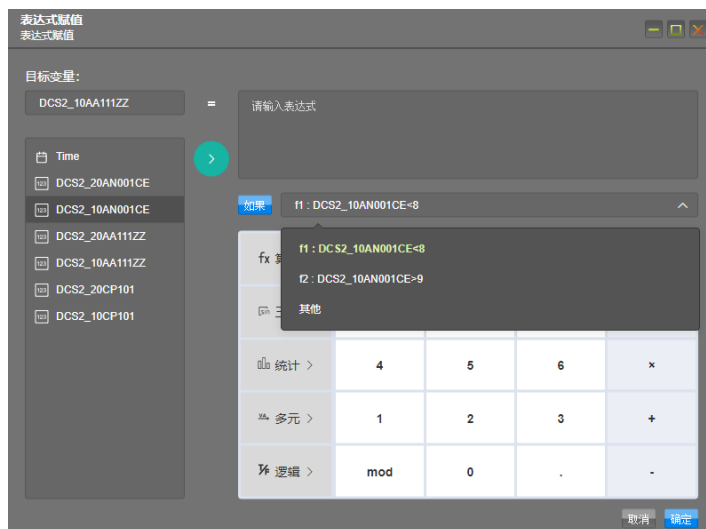


图104 启用条件语句后的表达式赋值弹窗

4.6.6.3 平移赋值

当目标变量是由某个原始变量的所有单元格沿行序号向前或向后平移生成时，可以使用平移赋值功能。平移赋值需要用户指定四方面内容：

- 目标变量：即需要生成的列；
- 原始变量：即作为参考的原始变量；
- 平移方向：向上（沿行序号减小方向）或向下（沿行序号增大方向）；
- 平移量：沿平移方向的移动行数。

提示：由平移赋值生成的列不会改变原始表格的行数，首尾位置会对应平移方向和平移量置空或截断。



图105 平移赋值



4.6.6.4 采样赋值

采样赋值用于快速生成用户指定规则、范围和样本大小的数据集。采样赋值的具体参数说明见 5.1.2 节“数据采样”。

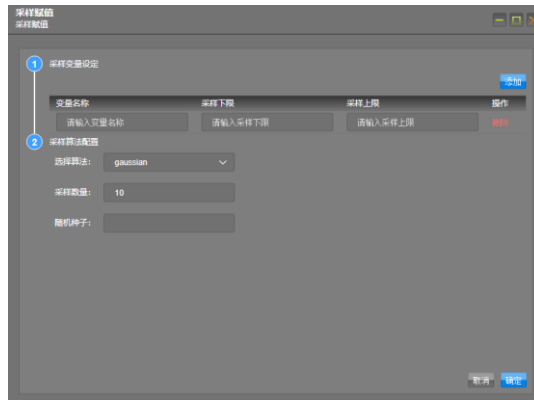


图106 采样赋值

4.6.7 数据合并

数据合并功能提供了将多个数据集合并为单一数据集的功能，且具体分为两类合并方式，即合并样本（纵向拼接）与合并变量（横向拼接）。数据合并功能的入口为左侧资源树-数据管理中数据集的右键菜单选项，如图 107 所示。

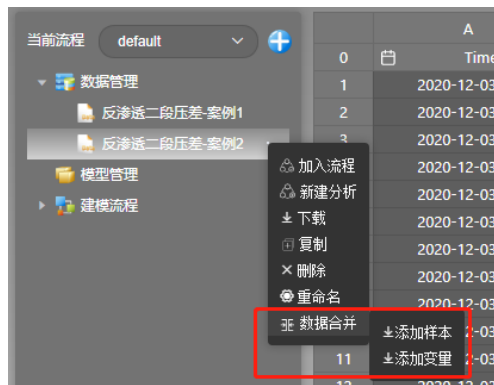


图107 数据合并

4.6.7.1 选择合并数据集

完成数据合并需要首先确定 2 个数据集：原始数据集 A，以及作为合并素材的数据集 B。其中，数据集 A 由资源树上的选中入口确定，数据集 B 则由数据合并弹窗中选择，如图 108 所示。选择方式有 2 种：一是从资源树的数据管理列表中直接选择，也可以从本地读取 dttable 文件进行选择。





图108 选择合并数据集

4.6.7.2 添加样本

当原始数据集需要补充更多的样本（即增加行数）时，可以使用添加样本功能。合并后的新数据集表格结构可由用户指定，如图 109 所示：

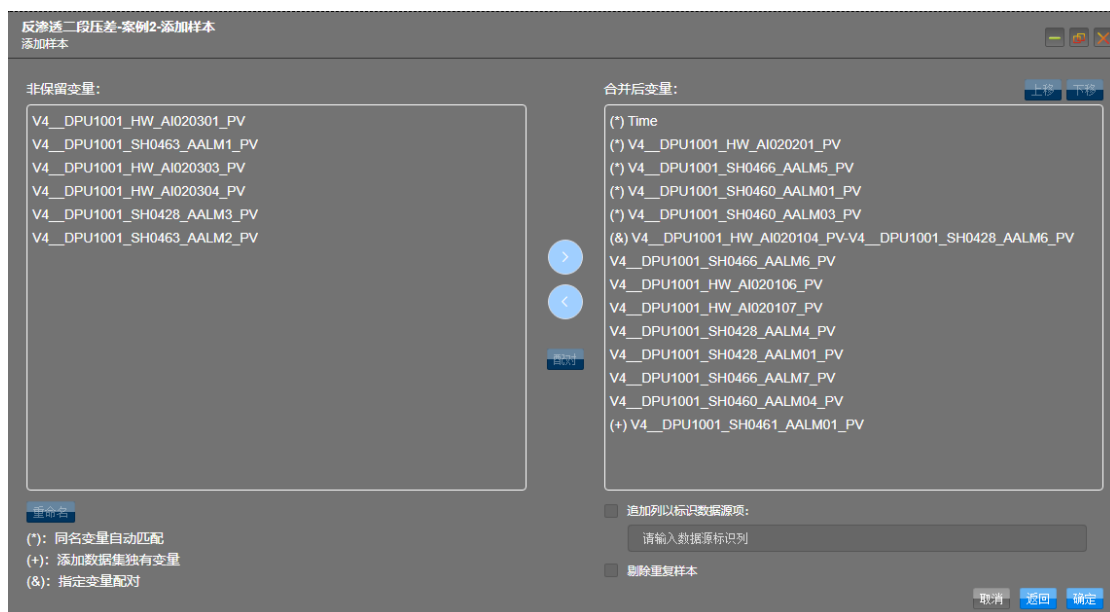


图109 添加样本

- 原始数据集 A 中的独有变量，见“合并后变量”中没有特殊符号标注的变量；
- 原始数据集 A 和素材数据集 B 的同名公共变量，会默认自动拼接为同一列，并在“合并后变量”中用“(*)”标注；



- 素材数据集 B 中的独有变量，默认存在于左侧的“非保留变量”，可以选中后添加至右侧，并在“合并后变量”中用“(+)”标注，且支持重命名操作；
- 用户可以手动指定原始数据集 A 和素材数据集 B 中的 2 个变量可以合并为同一个变量，具体操作方式为分别选中左侧“非保留变量”中的单个变量，以及右侧“合并后变量”中的单个没有特殊符号标记的变量，之后点击画布中央的“配对”按钮，即可将二者视为相同的变量，成功配对后会用“(&)”标注；
- 如需取消已经实现的变量添加或配对，可以选中右侧中的对应变量添加回左侧；
- 勾选“追加列以标识数据源项”并填入列名，即可在合并后数据集中显示当前行样本的数据来源；
- 勾选“剔除重复样本”，会将两个数据集中完全相同的样本仅保留一份。

4.6.7.3 添加变量

当原始数据集需要补充更多的变量特征（即增加列数）时，可以使用添加变量功能。合并后的新数据集表格结构可由用户指定，如图 110 所示：



图110 添加变量

- 从左侧“非保留变量”中选择一个或多个变量添加至右侧，添加后的变量用“(+)”标注，且支持重命名操作；



- 可以从左侧“非保留变量”和右侧“合并后变量”中各选择一个没有特殊符号标记的变量，并点击画布中的“配对”按钮，将二者视为相同的变量，配对后的变量用“(&)”标注，并将开启“样本匹配”功能；支持有多个配对变量；
- “样本匹配”功能开启需保证有至少一组已经配对的变量，开启后将根据匹配规则完成样本行的合并功能；四种匹配规则分别为“仅保留相同样本”“保留当前数据集样本”“保留添加数据集样本”以及“保留全部样本”；可以选择是否对合并后的样本进行排序；
- 勾选“追加列以标识数据源项”并填入列名，即可在合并后数据集中显示当前行样本的数据来源。

4.7 数据可视化

DTEmpower 支持依托“数据处理”菜单对工程中的数据绘制包括 1D 箱线图、2D 散点图、3D 散点图、高维可视化、相关图、平行图、柱状图、折线图在内的可视化图表，用以辅助用户对数据集当前情况进行查看和分析。

4.7.1 1D 箱线图绘制

单击顶部数据处理菜单中的 1D 箱线图按钮，在弹出窗口中选择变量（支持同时选择多个绘制变量），单击确定即可进行 1D 箱线图的绘制，绘图结果如图 111 所示。

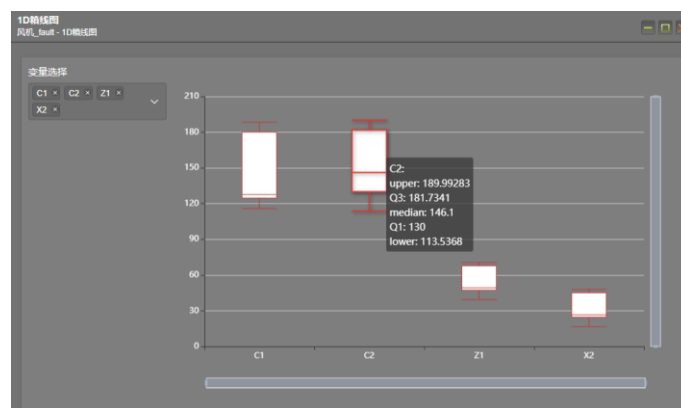


图111 1D 箱线图绘制



鼠标悬浮至某个变量的箱图上方，即可查看其具体信息，从上至下分别为：变量名、除去异常值外的最大值（upper）、75%百分位数（Q3）、中位数、25%百分位数（Q1）、除去异常值外的最小值（lower）。

4.7.2 2D 散点图绘制

单击顶部数据处理菜单中的 2D 散点图按钮，在弹出窗口中选择变量 X 轴变量和 Y 轴变量后，单击确定即可进行 2D 散点图的绘制，如图 112 所示。

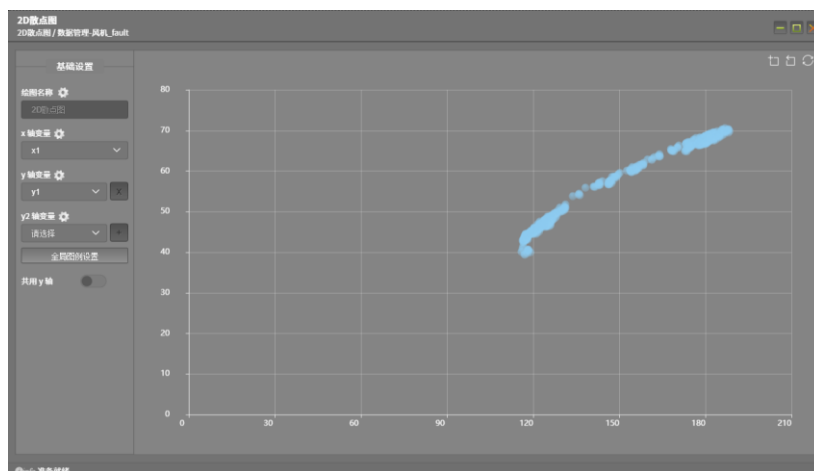


图112 2D 散点图绘制

当用户指定多个 Y 轴变量时，可以选择是否共用相同的 Y 轴。共用 Y 轴和非共用 Y 轴绘制的 2D 散点图分别如图 113、图 114 所示。

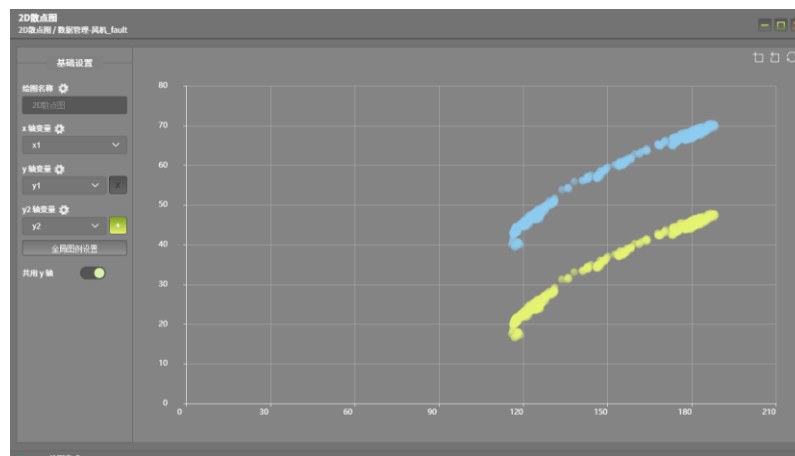


图113 2D 散点图绘制（共用 Y 轴）



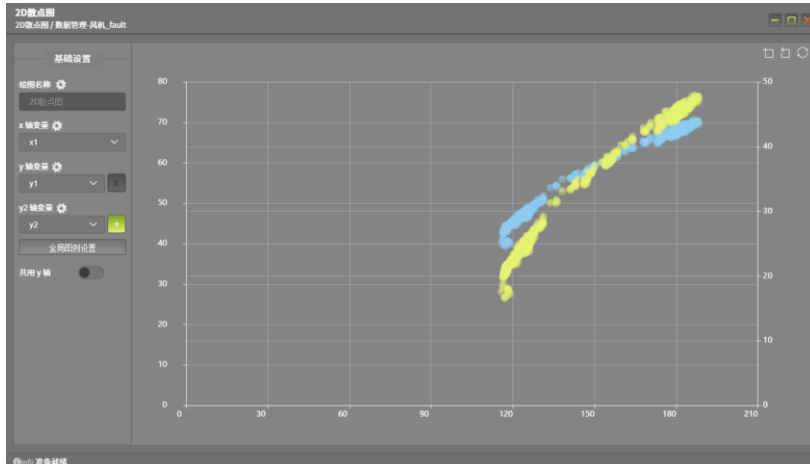


图114 2D 散点图绘制（非共用 Y 轴）

在 2D 散点图中，将鼠标悬浮至散点上可显示对应的变量名称与数值。

4.7.3 3D 散点图绘制

单击顶部数据处理菜单中的 3D 散点图按钮，在弹出窗口中选择变量 X 轴变量、Y 轴变量和 Z 轴变量后，单击确定即可进行 3D 散点图的绘制，如图 115 所示。

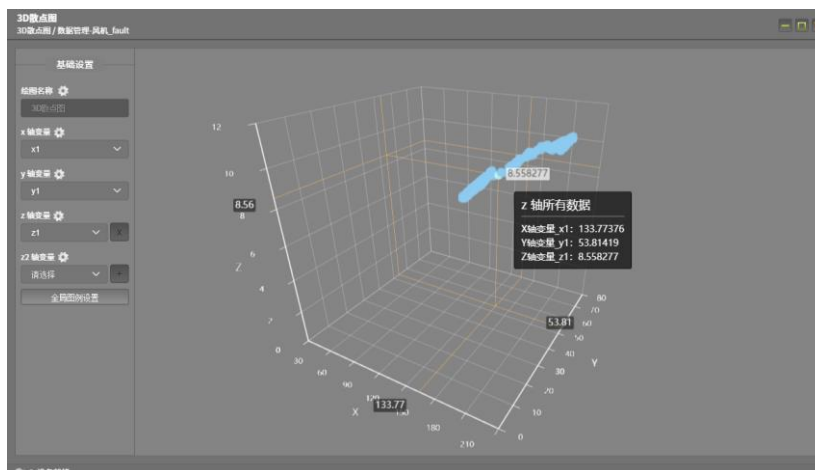


图115 3D 散点图绘制

在 3D 散点图中，将鼠标悬浮至散点上可显示对应的变量名称与数值。

4.7.4 高维可视化绘制

单击顶部数据处理菜单中的高维可视化按钮，在弹出窗口的绘制变量中选择多个变量，指定绘制维数，以及其他算法配置，如图 116 所示。



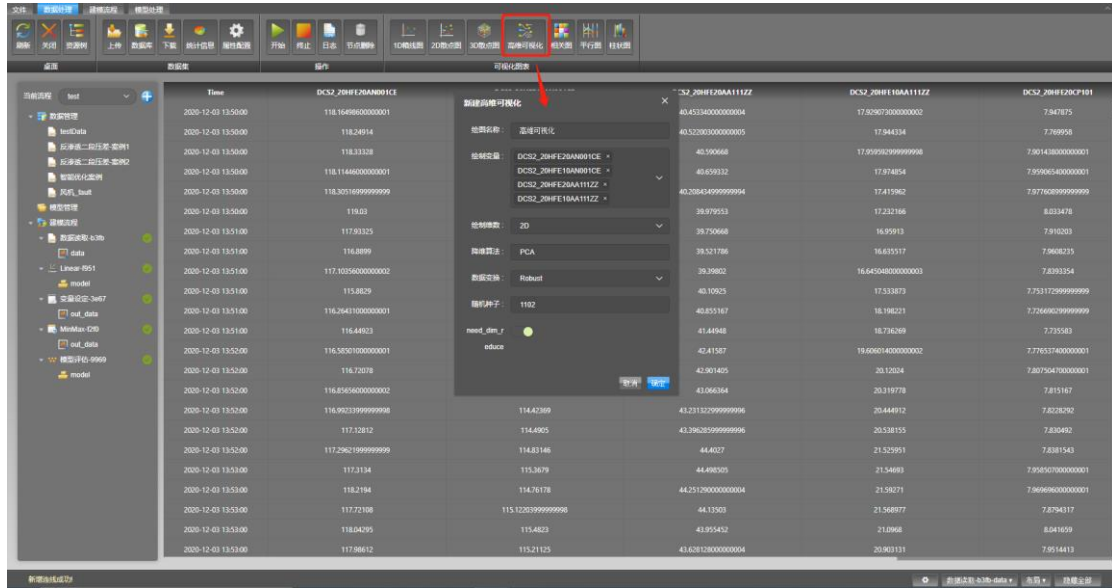


图116 高维可视化绘制

点击确定后，即可将多个维度的变量信息在可见维度下显示，从而观察数据的分布规律等。如图 117 所示为 4 维变量的可视化结果，通过降维算法将原始 4 维空间分布的数据展示在 2 维平面中，并在 y 轴变量设置中开启聚类分析，可以对数据的分布有更直观的认识。

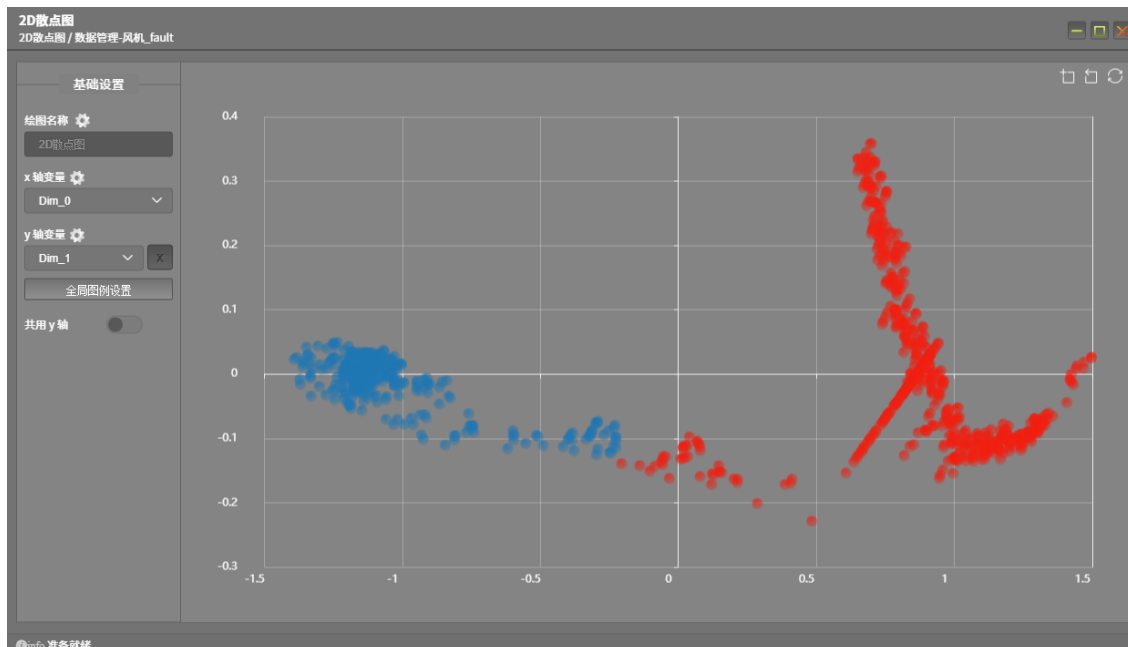


图117 高维可视化绘图结果



4.7.5 相关图绘制

单击顶部数据处理菜单中的相关图按钮，在弹出窗口中选择需要进行相关性分析的变量（支持复选）后，单击确定即可进行相关图的绘制，如图 118 所示。

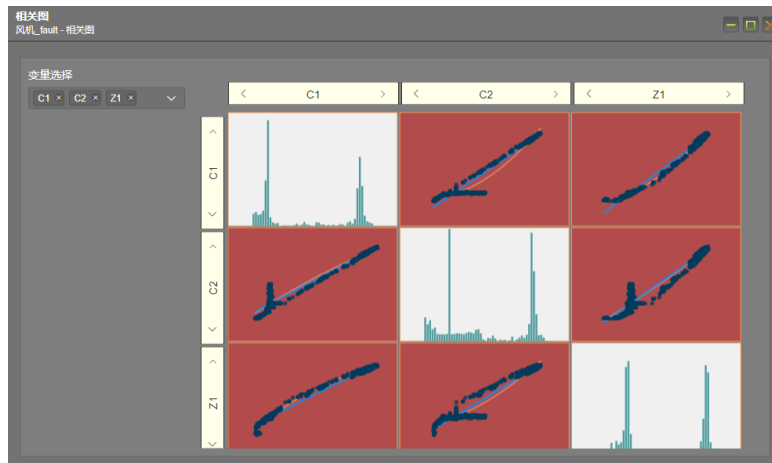


图118 相关图绘制

在相关图中，主对角线上为对应变量的分布直方图，其余为两个变量之间的散点图，同时绘制了变量之间的线性拟合函数曲线与多项式拟合函数曲线。

4.7.6 平行图绘制

单击顶部数据处理菜单中的平行图按钮，在弹出窗口中选择需要进行平行图绘制的变量（支持复选，且至少 2 个）后，单击确定即可进行平行图的绘制，如图 119 所示。

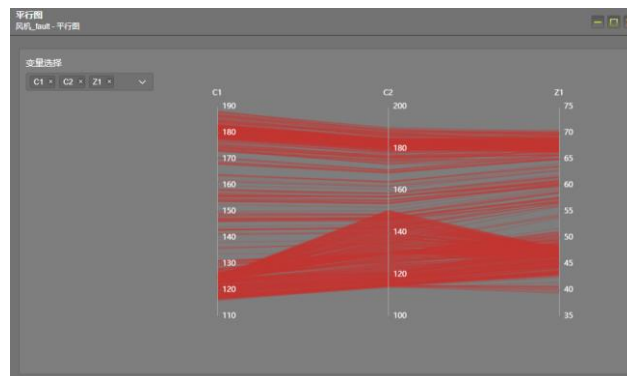


图119 平行图绘制



4.7.7 柱状图绘制

单击顶部数据处理菜单中的柱状图按钮，在弹出窗口中选择需要进行柱状图绘制的变量（支持复选）后，单击确定即可进行柱状图的绘制，如图 120 所示。

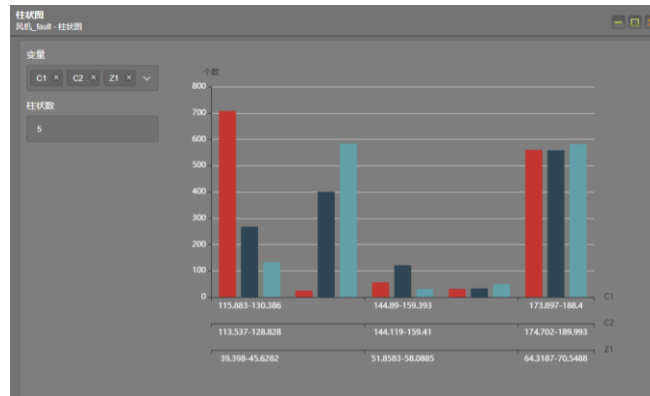


图120 柱状图绘制

4.7.8 折线图绘制

单击顶部数据处理菜单中的折线图按钮，在弹出窗口中选择需要进行折线图绘制的变量（仅支持单个变量）后，单击确定即可进行折线图的绘制，如图 121 所示。

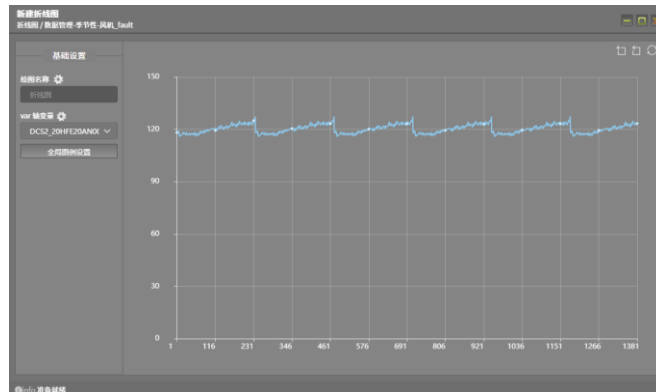


图121 折线图绘制

4.8 数据可视化交互配置

DTEmpower 支持用户对可视化数据图表元素进行自定义配置，如图标题、坐标轴、图例等，使绘制的图表更符合用户的需要。



4.8.1 标题设置

进入可视化图形后，点击绘图名称右侧的齿轮，即可打开标题设置，如图 122 所示。

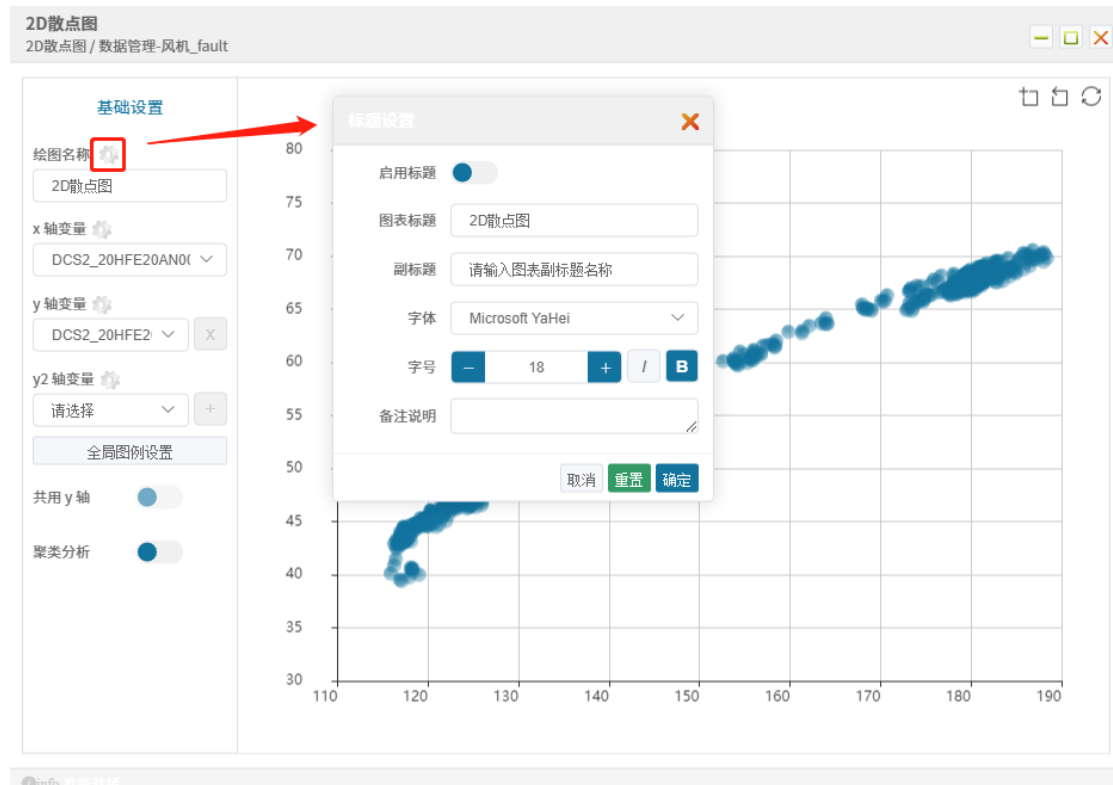


图122 标题设置

各设置描述如下：

- 启用标题：开启后将在图表上方显示标题，默认为关闭状态；
- 图表标题：设置标题文字；
- 副标题：可选项，设置副标题文字；
- 字体：设置标题字体；
- 字号：设置标题字号大小、斜体、加粗等；
- 备注说明：可选项，用以记录当前配置的描述信息。

4.8.2 x 轴变量设置

进入可视化图形后，点击 x 轴变量右侧的齿轮，即可打开 x 轴变量设置，如图 123 所示。

以 2D 散点图为例，x 轴变量设置包括范围设置与标签设置两类。





图123 x 轴变量设置

各设置描述如下：

- 自适应：开启后，坐标轴区间将自动根据对应的数值范围进行适配；
- 是否归一：开启后，根据原始数值范围进行缩放，将坐标轴区间固定为 0-1；
- 最小值：未开启自适应前提下，用户可手动设定坐标轴最小值；
- 最大值：未开启自适应前提下，用户可手动设定坐标轴最大值；
- 刻度：未开启自适应前提下，用户可手动设定网格的个数；
- 显示标签：开启后，将在坐标轴附近显示轴标签，默认为关闭状态；
- 标签名称：设置轴标签文字；
- 字体：设置轴标签字体；
- 字号：设置轴标签字号大小、斜体、加粗等。

4.8.3 y/z 轴变量设置

进入可视化图形后，点击 y 轴变量右侧的齿轮，即可打开 y 轴变量设置，如图 124 所示。



以 2D 散点图为例，y 轴变量设置包括颜色设置、绘点设置、范围设置与标签设置四类。

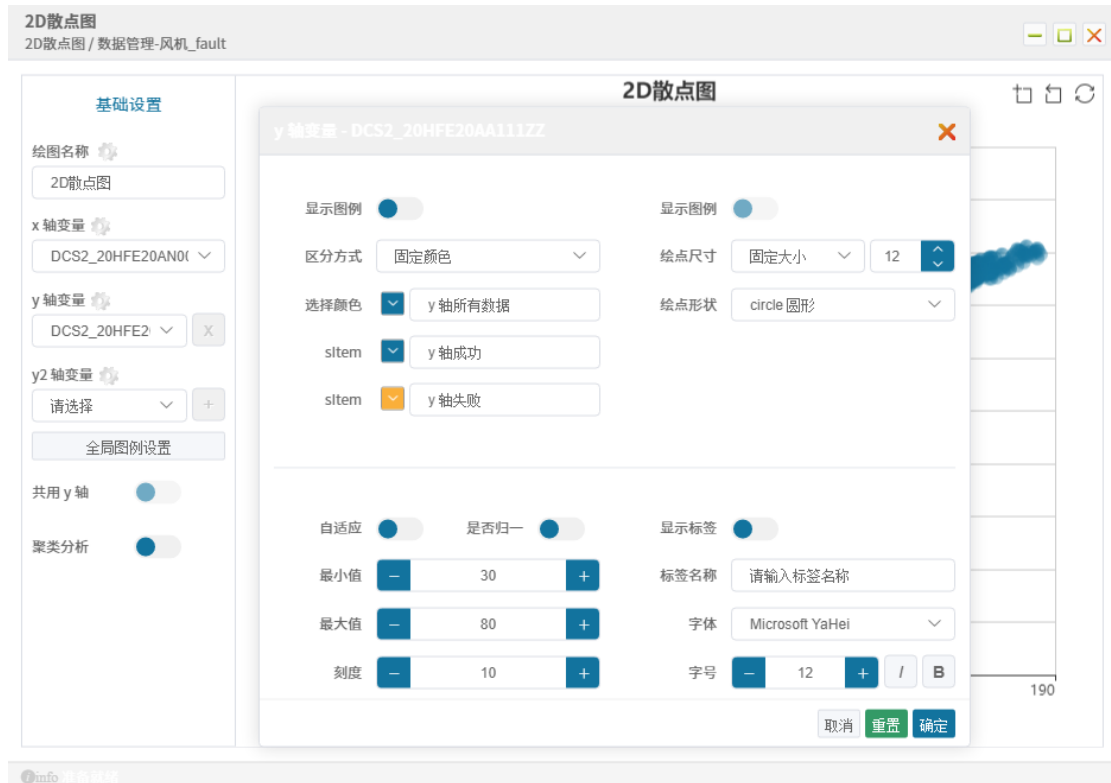


图124 y 轴变量设置

各设置描述如下：

- 显示图例：开启后，将在用户指定区域显示对应图例，默认为关闭状态；
- 选择颜色：设置对应变量的颜色，点击后会弹出颜色选择器；
- 绘点尺寸：（散点图专用）设置绘点的像素尺寸；
- 绘点形状：（散点图专用）设置绘点的形状；
- 自适应：开启后，坐标轴区间将自动根据对应的数值范围进行适配；
- 是否归一：开启后，根据原始数值范围进行缩放，将坐标轴区间固定为 0-1；
- 最小值：未开启自适应前提下，用户可手动设定坐标轴最小值；
- 最大值：未开启自适应前提下，用户可手动设定坐标轴最大值；
- 刻度：未开启自适应前提下，用户可手动设定网格的个数；
- 显示标签：开启后，将在坐标轴附近显示轴标签，默认为关闭状态；



- 标签名称：设置轴标签文字；
- 字体：设置轴标签字体；
- 字号：设置轴标签字号大小、斜体、加粗等。

4.8.4 新增副 y/z 轴

部分图形（如散点图）支持同时绘制多个 y 轴。通过点击最后一个 y 轴右侧的“+”号，即可新增一个副 y 轴，如图 125 所示。

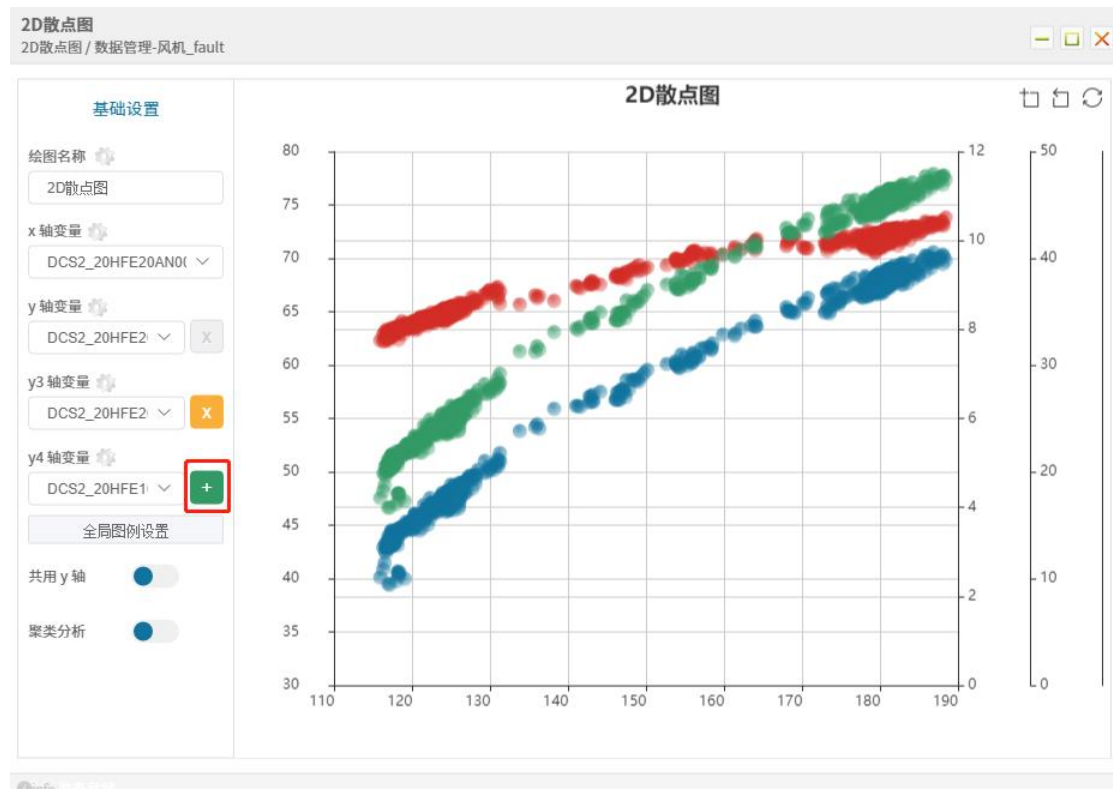


图125 新增副 y 轴

4.8.5 删除副 y/z 轴

对于多余的副 y 轴，可以通过点击对应变量的“x”号，实现该副 y 轴的删除，如图 126 所示。



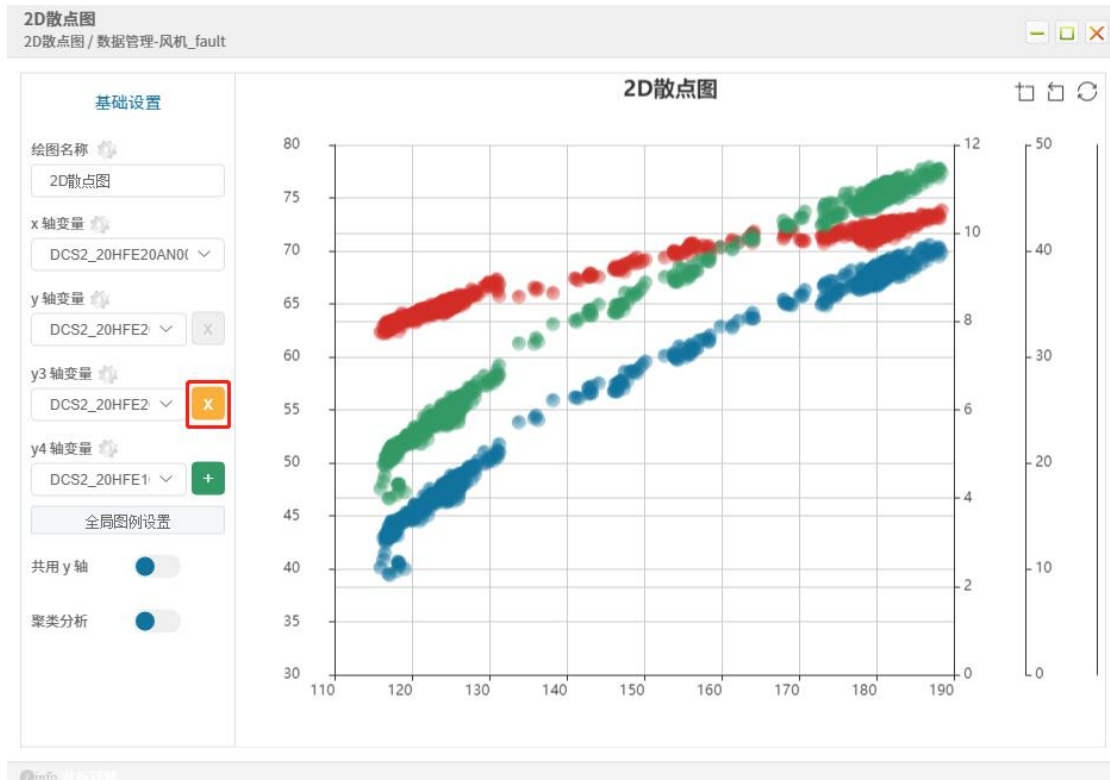


图126 删除副 y 轴

4.8.6 共用 y 轴

当存在副 y 轴时，用户既可以单独设定每个 y 轴的坐标区间，此时应使共用 y 轴按钮处于关闭状态，如图 127 所示；也可以让所有 y 轴变量共用一套坐标区间，此时应使共用 y 轴按钮处于开启状态，如图 128 所示。

注：共用 y 轴按钮默认置于关闭状态。



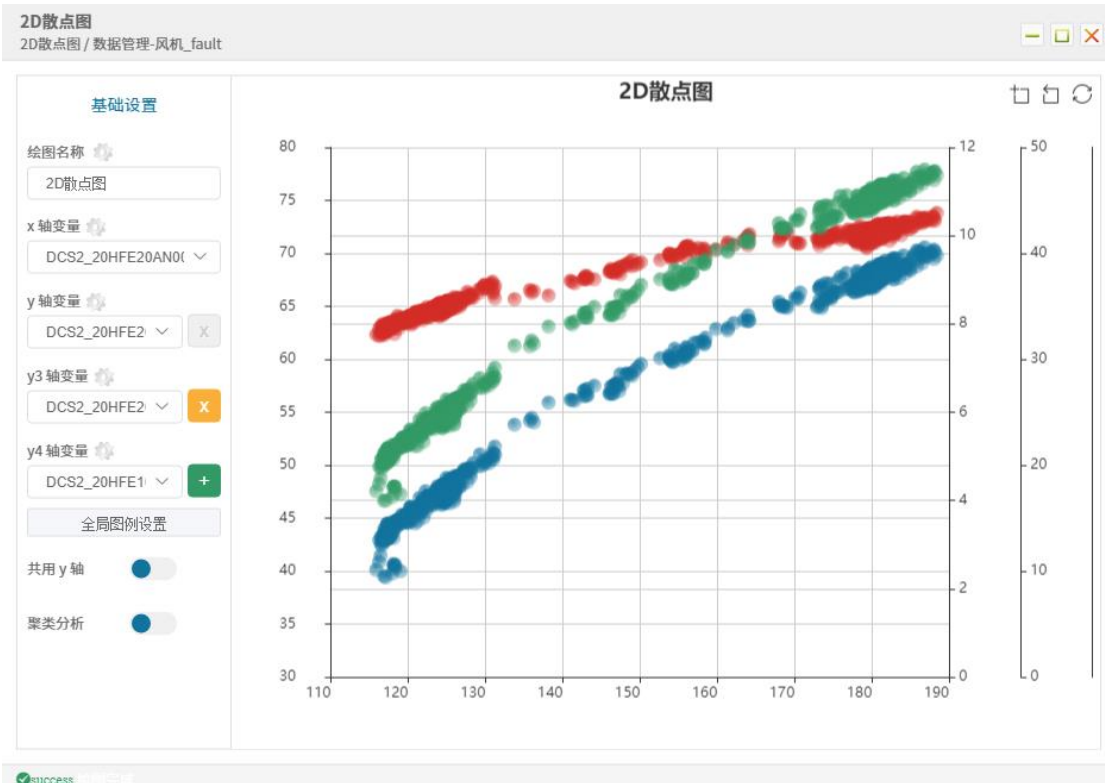


图127 共用 y 轴-关闭

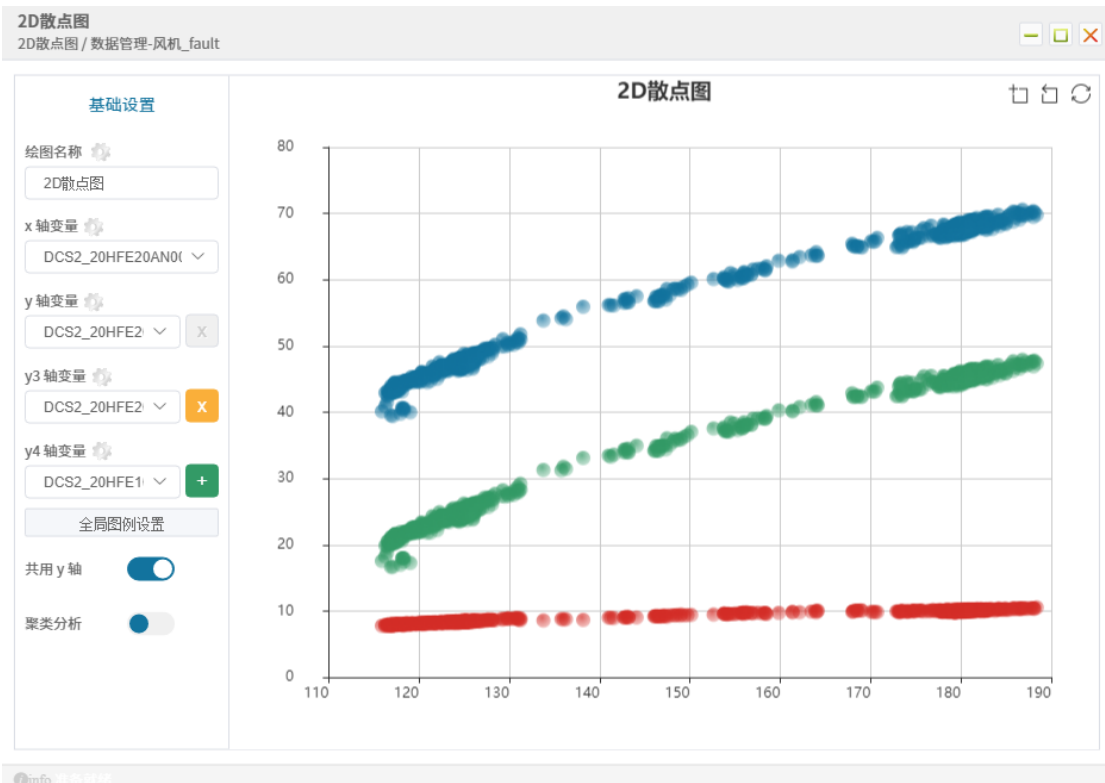


图128 共用 y 轴-开启



4.8.7 图例设置

进入可视化图形后，点击“全局图例设置”按钮，即可打开图例设置，如图 129 所示。

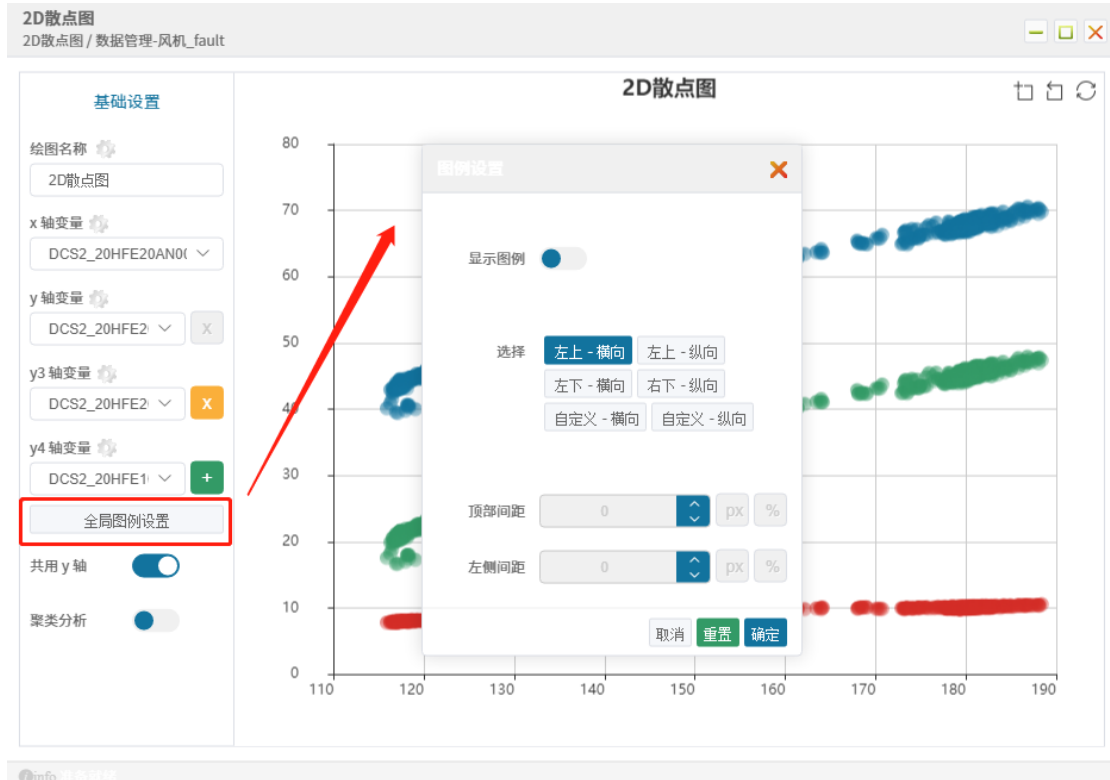


图129 图例设置

各设置描述如下：

- 显示图例：开启后将显示所有图例；
- 固定位置：设置图例所在位置；
- 自定义位置：当固定位置选择“自定义-横向”或“自定义-纵向”时开启，用户可通过设定顶部间距或左侧间距来改变图例所在位置。

4.8.8 区域缩放与还原

在可视化图形右上角提供三个按钮，从左至右分别为区域缩放、区域缩放还原与还原，如图 130 所示。



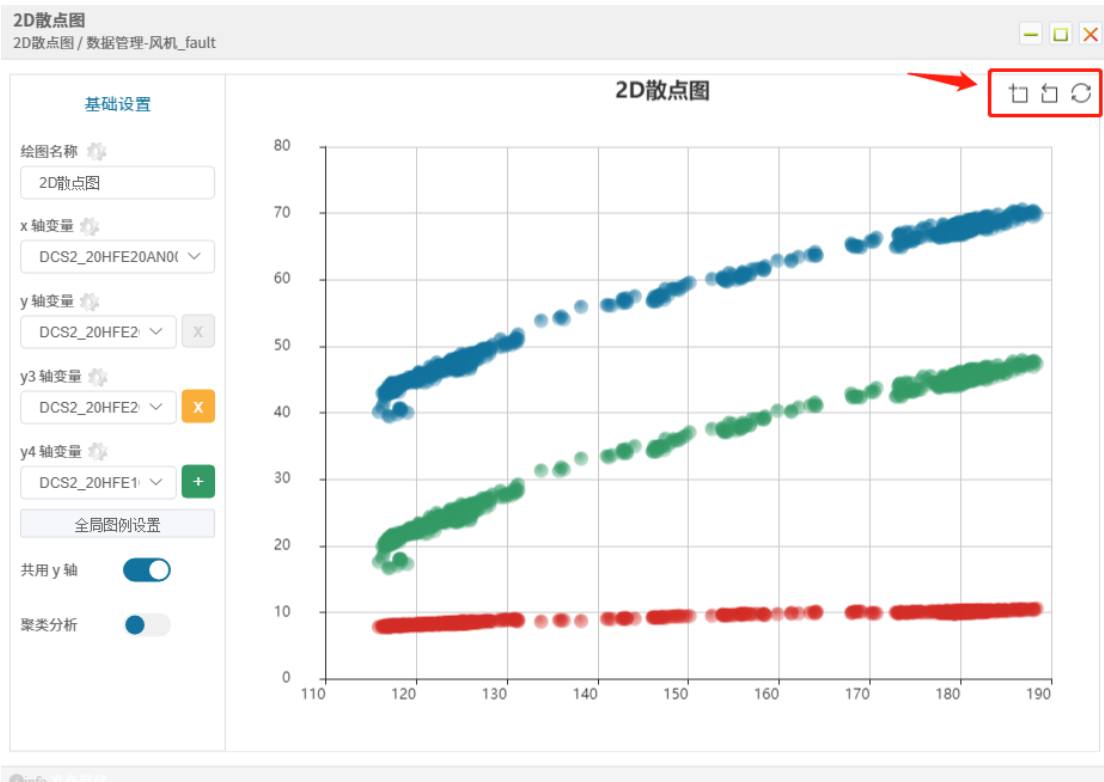


图130 区域缩放与还原

各按钮功能描述如下：

- 区域缩放：点击后，光标改变为“+”型，通过框选某矩形范围，可自动调整各坐标轴区间，使得最新视图与框选范围一致；
- 区域缩放还原：点击后，可撤销上一步区域缩放操作，并支持连续撤销；
- 还原：取消全部区域缩放操作，将图形还原为原始形态。

4.8.9 数据聚类可视化

在 2D 散点图或 3D 散点图中，可通过 y 轴变量设置中的聚类分析的下拉选项完成可视化，如图 131 所示。



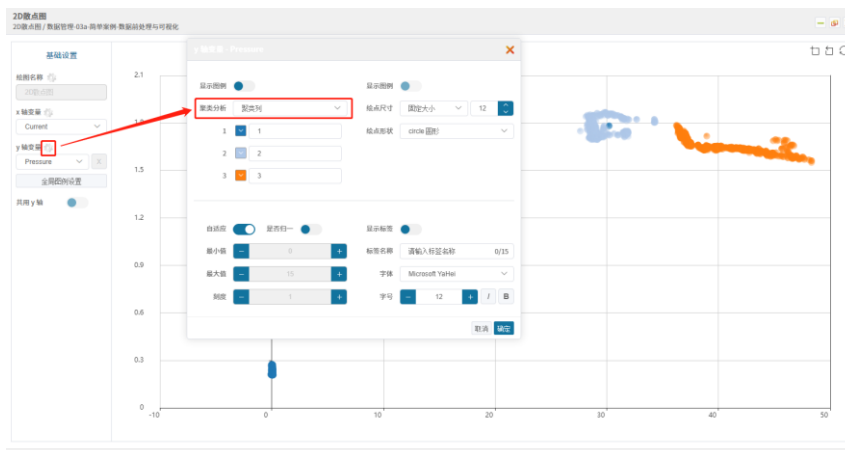


图131 数据聚类可视化

在聚类分析下拉选项中，除默认的固定颜色外，还可识别数据集中存在的有序变量或无序变量以作为聚类标识列。当选择了此类标识变量后，即可区分每个散点对应的类别，并通过不同的颜色（即分类标识）进行区分。用户可以自由更改标识对应的颜色。

应注意在数据聚类可视化操作下，每个坐标轴仅能选择一种变量（即不支持新增 Y 轴，否则聚类分析可选项默认关闭），且标识变量的类别数不超过 20 种。

4.9 数据源管理

DTEmpower 支持对数据库的连接功能。数据源管理模块提供了简单快捷的从数据库中提取数据集的方式，帮助用户实现“每一天的工作从连接数据源开始”。

需要说明的是，数据源管理功能非标准版 DTEmpower 搭载的功能。用户如需开放该功能，详情请咨询南京天沚软件有限公司。

4.9.1 连接数据源

连接数据源的入口有 2 处，分别为顶部 Ribbon 菜单的“数据库”按钮与左侧资源树的数据源管理右键菜单“连接”，如图 132 所示。需要注意的是，首次使用连接数据源功能的用户只能通过第一种方式进行连接，只有完成了常用数据源的配置后才能使用第二种连接方式。





图132 连接数据源

对于首次连接数据源的用户，需要补全相关信息后才能完成连接操作。点击顶部 Ribbon 菜单的“数据库”按钮后将出现相应弹窗，如图所示。在正确填写数据库类型、服务器 IP、端口号、用户名、密码后，点击底部“测试”按钮检查是否能正常连接至数据库。如连接成功，则顶部会出现“测试成功”提示字样，且数据库与数据表会变为下拉列表框以支持选择，如图 134 所示。如连接失败，会弹窗提示失败原因。另外，为避免连接响应时间过长，用户可通过设定“命令超时”的数值以判断是否提前终止连接。



图133 数据源配置



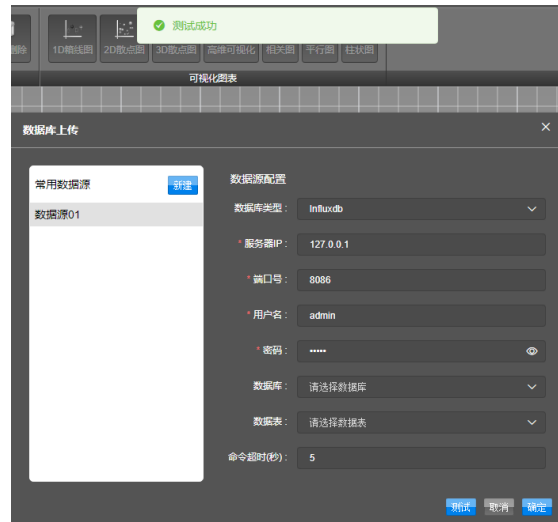


图134 成功连接至数据库

选择了所需要的数据库与数据表后，点击底部“确定”按钮可以进一步完成数据集结构的配置，如图 135 所示。用户可从左侧原始变量列表中勾选需要保存至数据集的变量添加至右侧。除了可对变量进行配置外，还支持对样本的配置功能。用户可以在导入全部、从前选择、以及基于时间稀释等多种方式中选择构建合适样本规模的数据集。

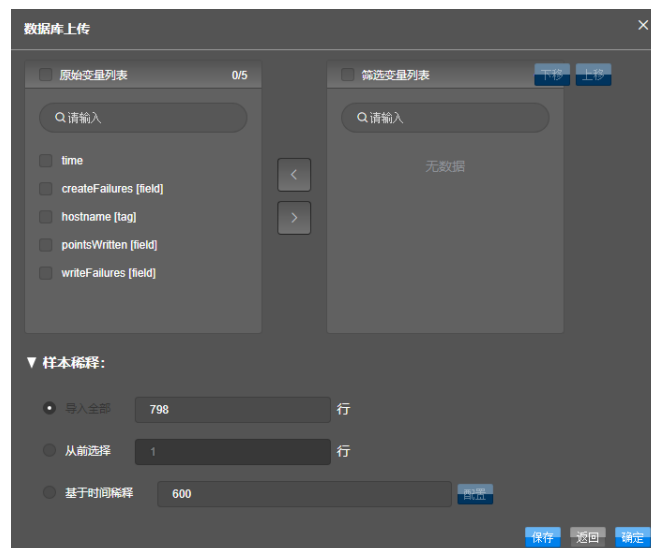


图135 连接数据源-数据集结构配置

当配置完毕后，通过点击底部“确定”按钮可完成连接数据源。提取的数据集将添加至左侧数据集的数据管理中。



4.9.2 新建数据源

新建数据源有 2 类入口：分别为通过顶部 Ribbon 菜单的“数据库”按钮进入配置弹窗中，点击左侧“新建”按钮；以及通过左侧资源树的数据源管理右键菜单中的“新建”选型。需要注意的是，左侧资源树的数据源管理层级默认是隐藏的，因此首次配置必须通过第一类入口实现。

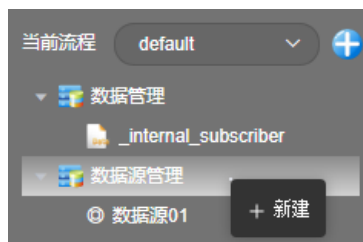


图136 新建数据源

4.9.3 保存数据源

当用户补全了常用数据源的配置信息后，可以通过“保存”按钮存储之前的配置信息，后续即可从资源树的数据源管理选项中直接使用该配置信息进行连接。

“保存”按钮见图 135 的底部菜单。

4.9.4 编辑数据源

当用户需要修改常用数据源保存的配置信息时，可以通过左侧资源树中数据源管理下右键菜单中的“编辑”选项进行修改，如图 137 所示。

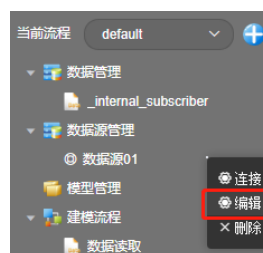


图137 编辑数据源

4.9.5 删除数据源

通过左侧资源树中数据源管理下右键菜单中的“删除”选项，可以删除对当前选中数据源的保存。



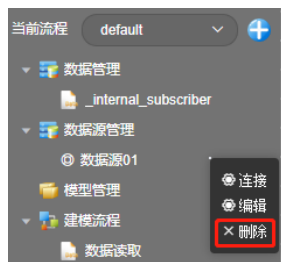


图138 删除数据源

4.10 建模流程

4.10.1 加入流程

在左侧资源树-数据管理中任意数据集上单击鼠标右键，在右键快捷菜单中选择“加入流程”，即可将对应数据集以数据载入节点的形式加入到当前建模流程中，如图 139 所示。

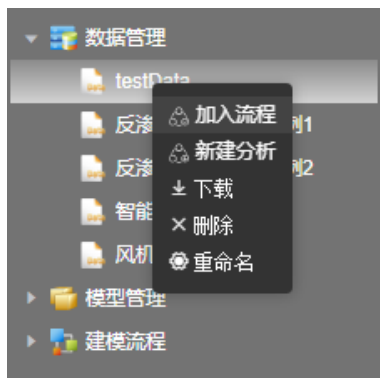


图139 加入建模流程

4.10.2 新建分析

在左侧资源树-数据管理中任意数据集上单击鼠标右键，在右键快捷菜单中选择“新建分析”，即可将对应数据集以数据载入节点的形式加入到新表格式建模流程中并运行，新流程默认名称与数据集名称一致（如当前流程不包含任何节点，则加入当前流程），如图 140 所示。



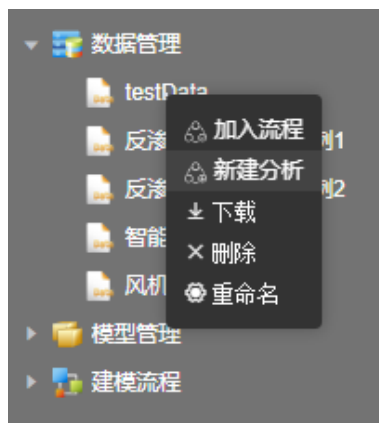


图140 新建分析流程

该新建流程选项将默认进入表格式建模流程，因此不会显示画布并隐藏左侧资源树。如需返回拖拽式建模流程，可在顶部菜单栏中点击“关闭”按钮。

4.10.3 添加节点

在工具箱中选中需要添加至建模流程中的功能节点后，按住鼠标左键不动，将节点拖拽至流程中，即可实现建模流程功能节点的添加，如图 141 所示。

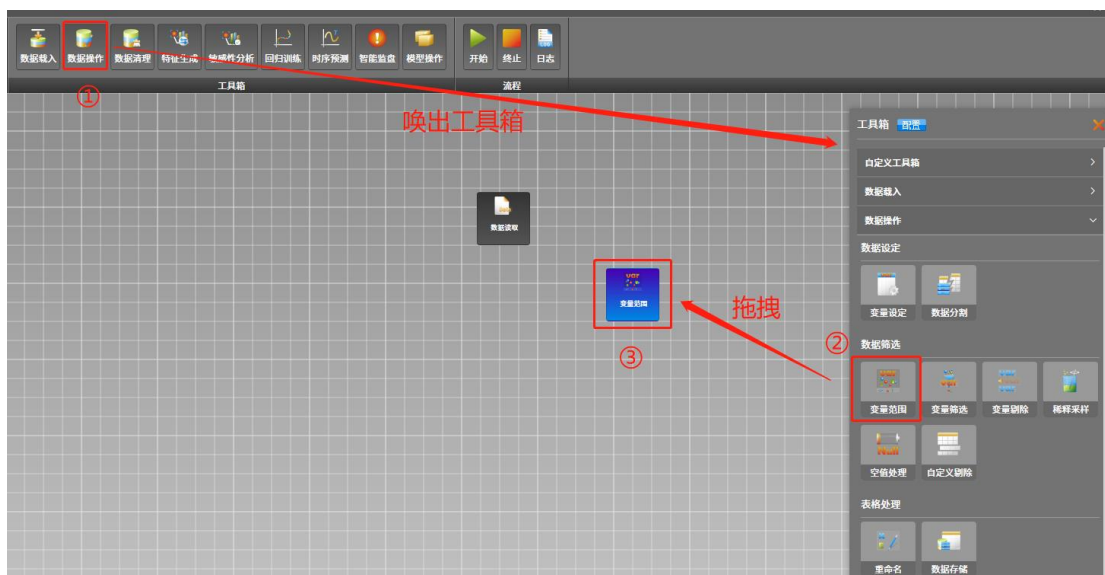


图141 添加节点

4.10.4 节点属性配置

鼠标左键单击节点；或鼠标右键单击节点，在右键快捷菜单中选择“查看”详情，即可在弹出窗口中对节点属性进行配置，如图 142 所示。



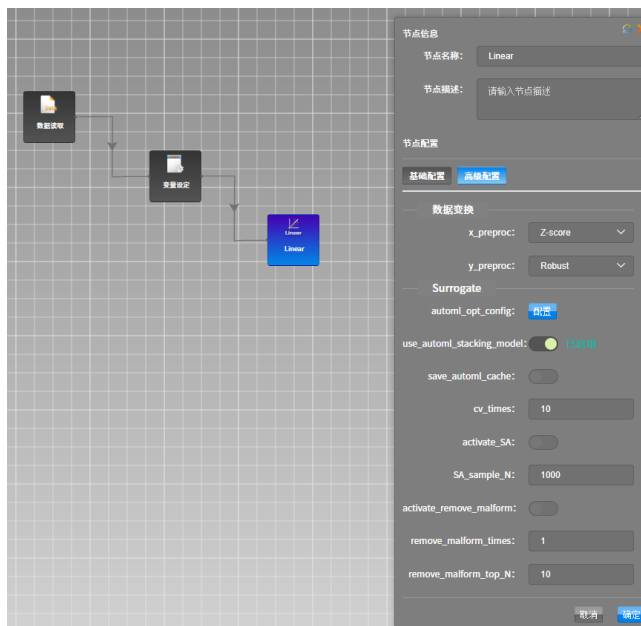


图142 节点属性配置弹窗

4.10.5 节点连线

节点间的连线除了代表节点的执行顺序外，通常还表征着数据在节点之间的传递关系。需要进行节点连线时，首先将鼠标移至上游节点右上角的连接点处如图 143 所示。



图143 连接点

在上游节点连接点处按住鼠标左键进行拖拽，将引线末端移动至下游节点即可完成节点的连接，如图 144 所示。



图144 节点连接

节点连接完成后，自动生成连线，且通过箭头关系表征节点执行的顺序和数据传递的关系，如图 145 所示。



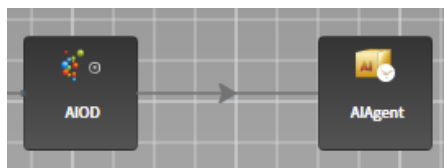


图145 节点连线

4.10.6 删除连线

在连线上方单击鼠标右键，在右键快捷菜单中选择删除连线，即可删除节点间的连接线，如图 146 所示。也可以在选中连线状态下通过 delete 快捷键完成删除。



图146 删除连线

4.10.7 节点间输入输出量关联

鼠标左键单击连线；或鼠标右键单击连线选择查看详情，即可在弹窗中对节点间传递的输入输出量进行关联，如图 147 所示。

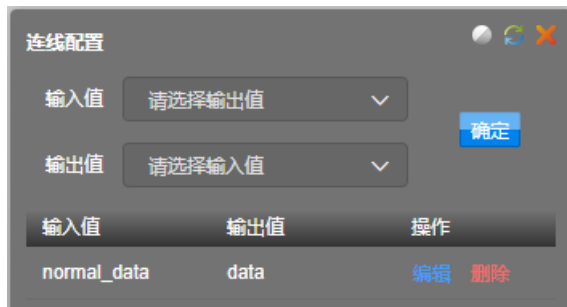


图147 连线配置

用户在选择输入值和输出值后单击确定按钮，即可完成节点间一对输入输出量的关联，当前连线已被关联输入输出将会出现在下方列表中，用户可通过编辑按钮修改已关联的输入输出量，也可通过删除按钮删除已关联的输入输出量。

应当注意的是，流程搭建过程中会根据上下游节点的输出与输入值类型完成连线自动配置。但对于部分具有复合输入或输出的节点，程序自动匹配的连线类型可能导致程序运行异常，此时需手动检查连线配置。



4.10.8 多选节点

在画布上按下键盘“Ctrl”键不放，同时按下鼠标左键拖拽矩形区域覆盖多个节点，即可实现多个控件的同时选中状态，如图 148 所示。

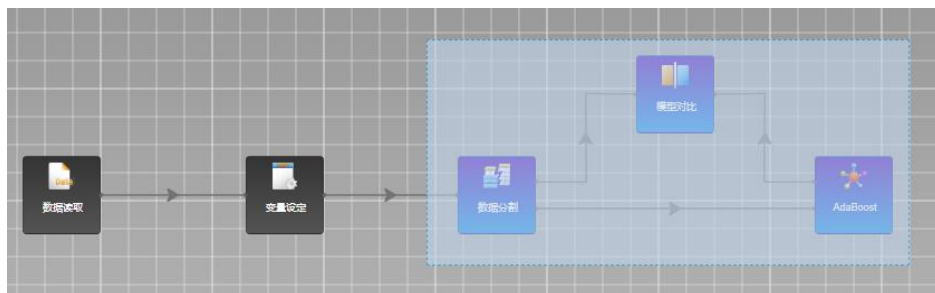


图148 多选节点

松开鼠标左键后，被选中节点会呈现高亮状态，同时自动弹出右键菜单操作选项，包括删除选中、水平对齐、垂直对齐、加入工具箱、水平等距对齐以及垂直等距对齐，如图 149 所示。

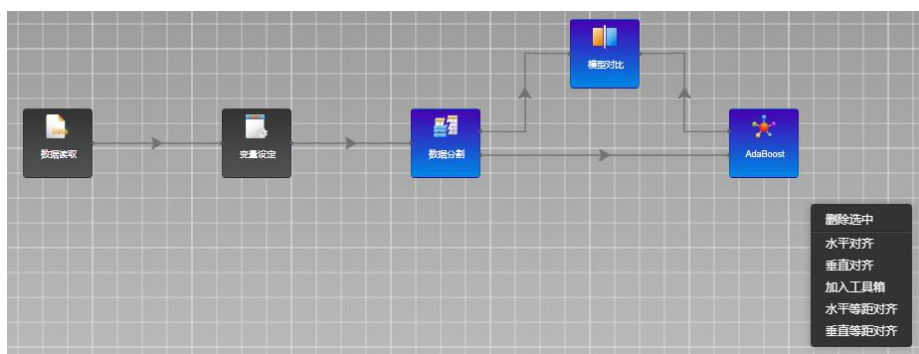


图149 多选节点右键菜单

4.10.9 删除节点

在节点上单击鼠标右键，在右键快捷菜单中选择删除节点即可完成节点的删除，如图 150 所示。也可以在选中节点状态下通过 delete 快捷键完成删除。



图150 删除节点



4.10.10 自定义工具箱

自定义工具箱为用户提供了收藏常用控件的功能，方便用户快速复用过去保存的节点并添加到新的流程中。自定义工具箱涉及三类步骤的操作：收藏、复用与配置。

4.10.10.1 收藏

在收藏操作中，用户需要在画布节点的右键菜单中选择加入收藏，并点击其中的一个自定义工具箱。成功后顶部会出现悬浮提示“添加收藏成功”字样。除可收藏单个节点，还支持多个节点及其连接线的收藏。无论是单节点收藏还是多节点收藏，自定义工具箱中均会保存其中的配置参数。另外需要注意的是，同一个自定义工具箱中不能出现同名控件，因此当重复添加同名节点时，系统会自动弹窗提示修改名称。



图151 自定义工具箱-收藏

4.10.10.2 复用

自定义工具箱中控件的复用方式与其他控件的使用方式基本一致。通过 Ribbon 菜单中的“工具箱”按钮唤出右侧工具箱后，即可见自定义工具箱下拉选项。选择已收藏的控件拖拽至画布中，即可复用之前配置好的控件。对于多选节点的收藏，会还原其原始位置。



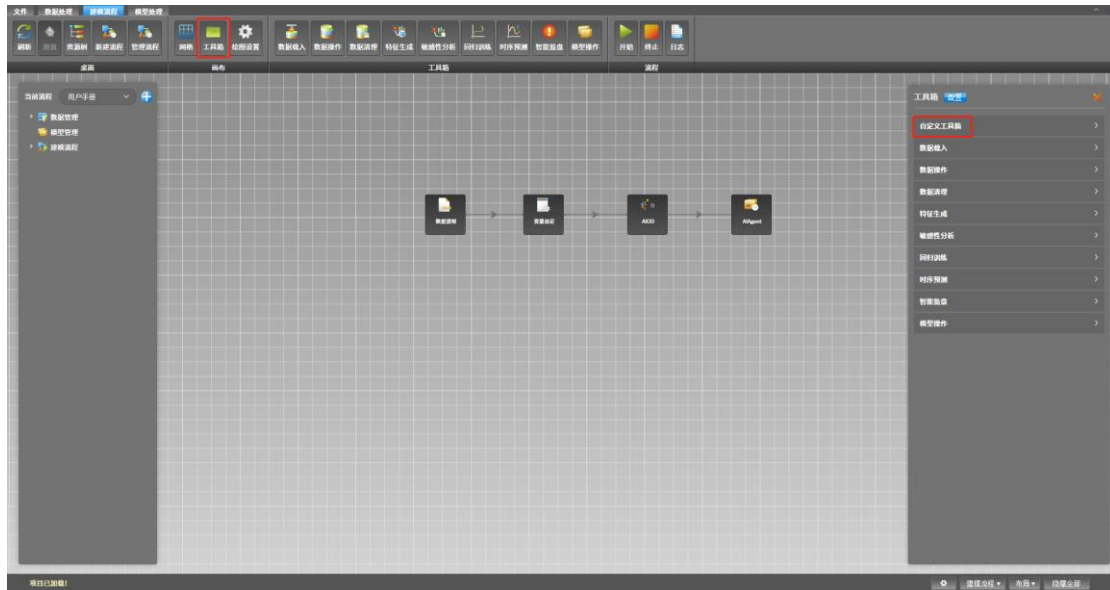


图152 自定义工具箱-复用

4.10.10.3 配置

在右侧工具箱顶部可见高亮的“配置”按钮，点击进入后即可见工具箱配置的相关操作选项，如图 153 所示。

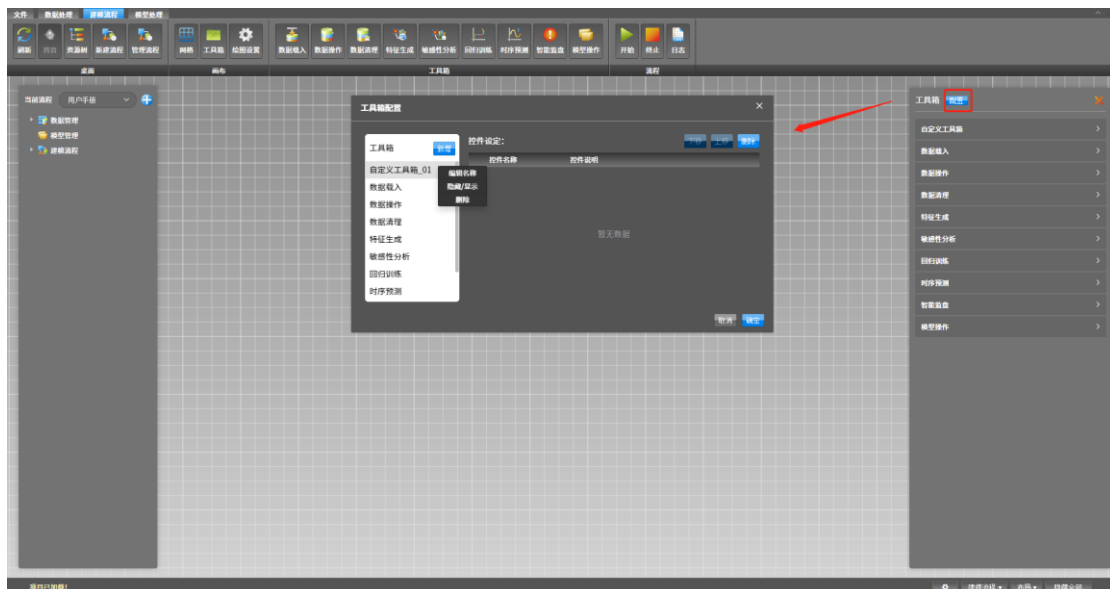


图153 自定义工具箱-配置

- 新增：增加自定义工具箱的个数；
- 编辑名称：右键菜单打开，修改选中自定义工具箱名称；
- 隐藏/显示：右键菜单打开，决定选中自定义工具箱在画布中是否可见；



- 删除（菜单）：右键菜单打开，删除选中自定义工具箱；
- 删除（按钮）：删除勾选状态的控件；
- 控件名称&控件说明：可编辑文本框，支持直接修改；
- 上移&下移：移动控件在自定义工具箱中的相对位置。

4.10.11 执行流程

建模流程搭建完成后，单击建模流程菜单中的开始按钮后，即可实现各节点的顺序执行，直至完成整个建模流程，如图 154 所示。

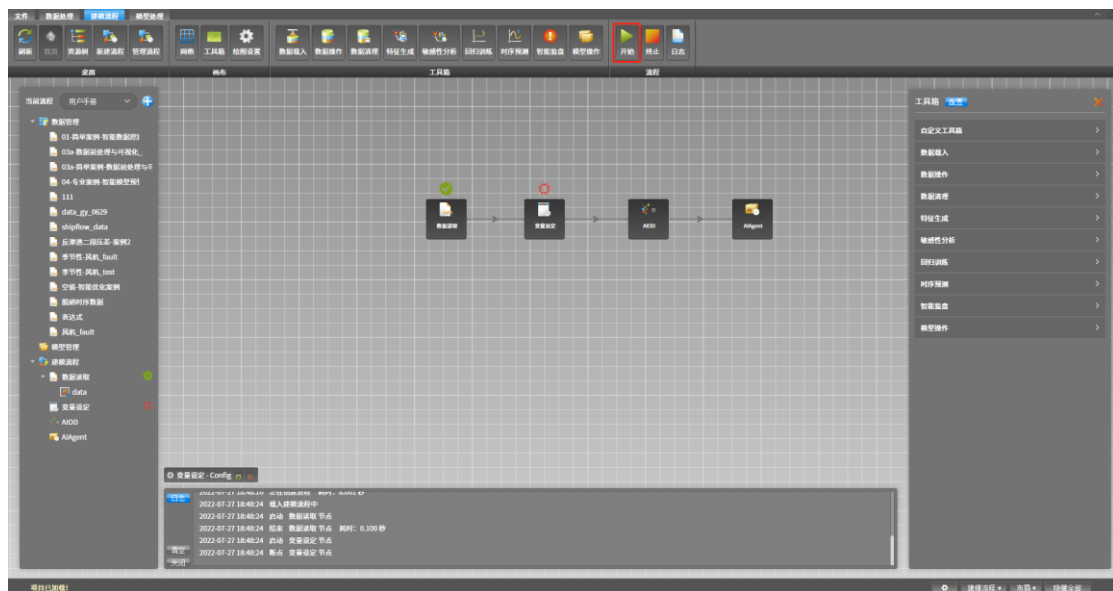


图154 执行流程

流程执行过程中，节点上方通过状态标志表征节点的运行状态，节点共分为运行中、运行报错、成功运行和等待用户进行交互式配置四种状态。同样的，左侧资源树中的建模流程下，会显示与画布中相同的节点运行状态。

4.10.12 运行日志查看/关闭

单击建模流程中的日志按钮，可对流程执行过程中的日志信息进行查看或隐藏，如图 155 所示。



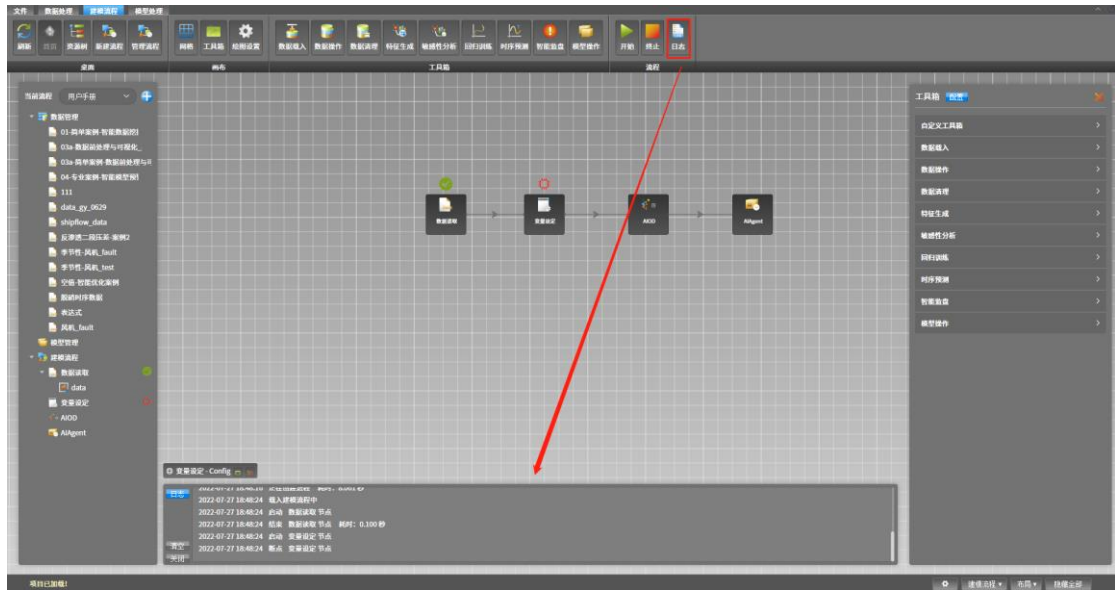


图155 流程运行日志

为了方便用户了解流程当前的执行状态，流程执行时日志窗口将自动展开。当日志窗口内容过多时，用户也可通过日志窗口左侧的清空按钮清空此前的运行日志信息。

4.10.13 新建流程

新建流程有两个操作入口，分别为顶部 Ribbon 菜单栏的“新建流程”按钮以及左侧资源树顶部的“⊕”按钮。点击按钮后会打开对应弹窗，如图 156 所示。

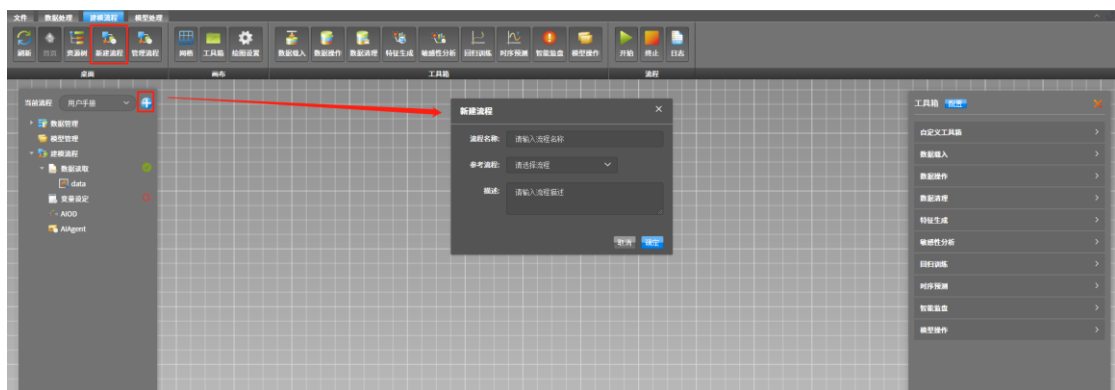


图156 新建流程

- 流程名称：必填项，用以标识新建的流程；点击确定后画布将自动跳转到新建立的流程；



- 参考流程：可选项，当选择了某个参考流程后，可以复制参考流程中的全部控件及其配置；
- 流程描述：可选项，用以记录该流程的描述信息。

4.10.14 管理流程

管理流程弹窗可以通过点击顶部 Ribbon 菜单栏中的“管理流程”按钮进入，如图 157 所示。

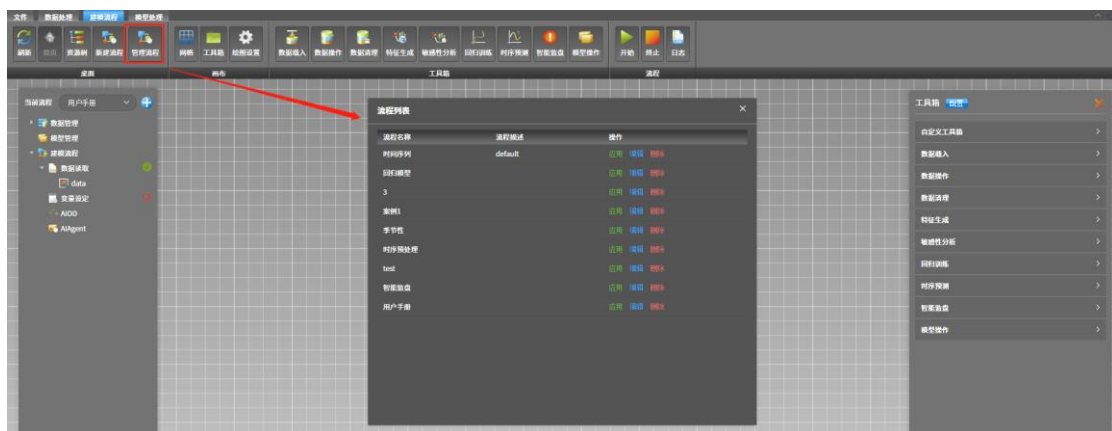


图157 管理流程

管理流程的功能如下：

- 应用：切换当前画布显示的流程；
- 编辑：编辑流程的名称与描述；
- 删除：删除选中流程。

4.10.15 终止流程

在流程执行过程中，用户可通过建模流程菜单中的终止按钮终止流程的执行，如图 158 所示。



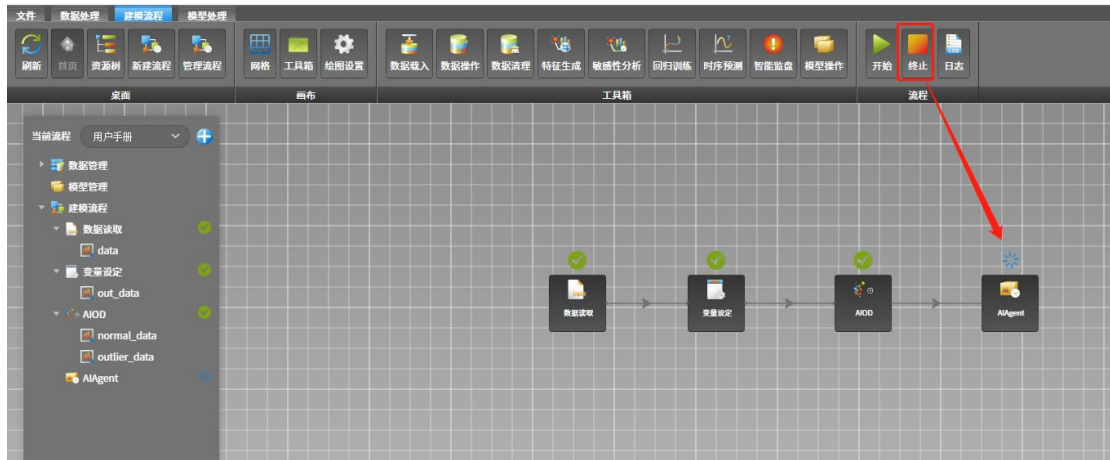


图158 终止流程执行

4.10.16 节点执行结果信息查看

节点执行完成后，单击左侧菜单树中的节点名称，节点会展开呈现执行结果信息，如图 159 所示。

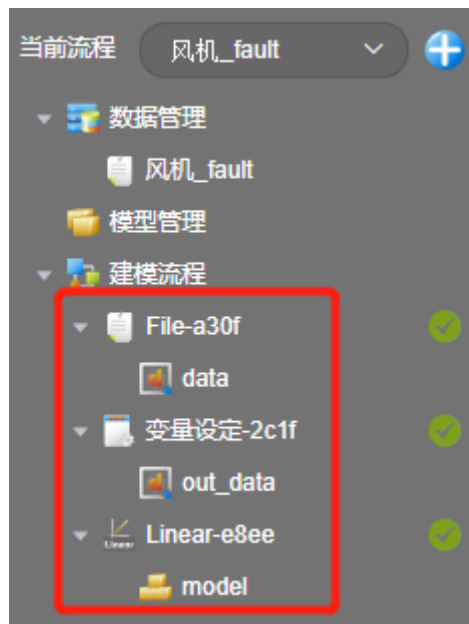


图159 节点执行信息菜单

在左侧资源树中单击任意节点执行结果信息，页面右侧即会对相应的结果信息进行展示，如图 160 所示。



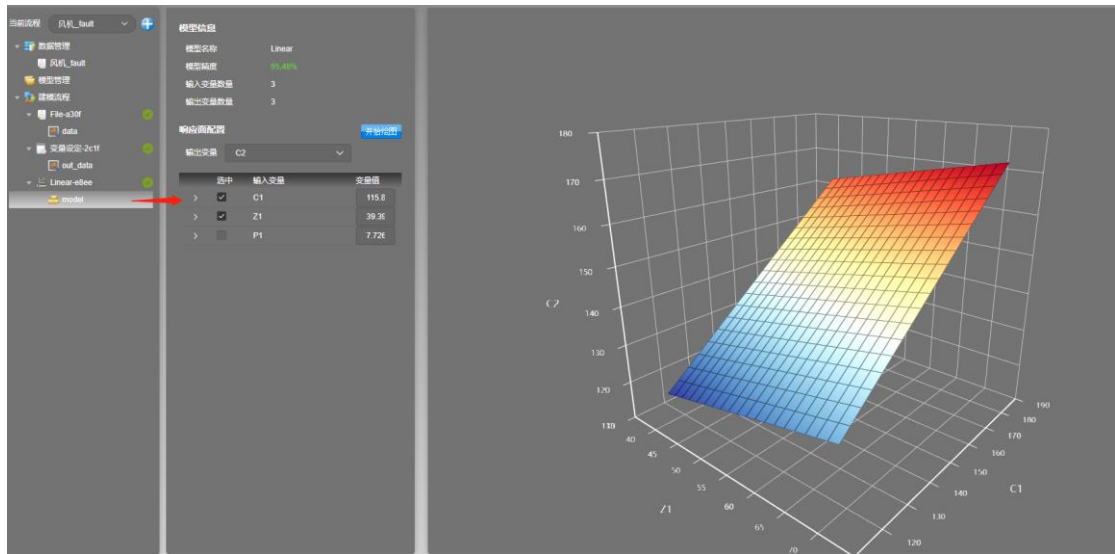


图160 节点执行信息查看

4.11 模型管理

4.11.1 保存模型

模型训练节点的执行结果即训练得到的模型，因此其资源树展开信息为训练好的模型，在模型上单击鼠标右键，选择保存，填写模型名称后即可将该模型保存至当前工程的模型管理中，方便用户进行模型的统一管理，如图 161 所示。

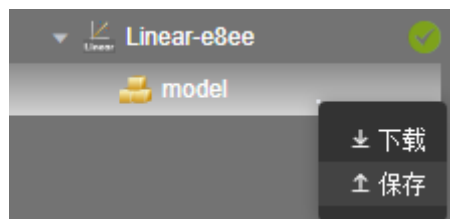


图161 保存模型

4.11.2 下载模型

对于左侧树状资源菜单中的模型（包括模型训练节点执行完成得到的模型和模型管理中存储的模型），用户在模型上单击鼠标右键，选择下载，即可将模型保存至本地磁盘中，实现模型跨平台的利用，如图 162 所示。





图162 下载模型

4.11.3 上传模型

在左侧资源树状菜单模型管理选项处单击鼠标右键，在右键快捷菜单中选择上传后在弹窗中选择存储在本地磁盘中的.dt 模型文件，即可将模型文件存储至当前工程的模型管理菜单下，如图 163 所示。

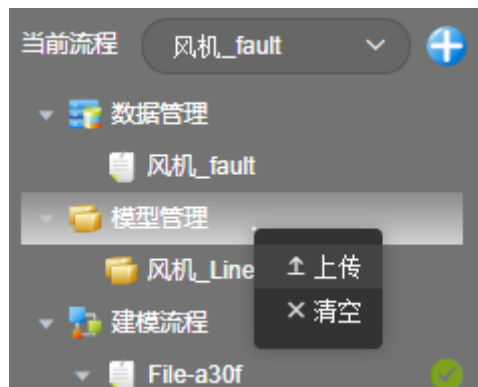


图163 上传模型

4.11.4 模型信息查看

对于左侧树状资源菜单中的模型（包括模型训练节点执行完成得到的模型和模型管理中存储的模型），在其上方单击鼠标左键，页面右侧即对该模型的响应面信息进行展示，用户依次选择输出变量和输入变量后单击开始绘图按钮即可查看该模型的响应面信息，如图 164 所示。



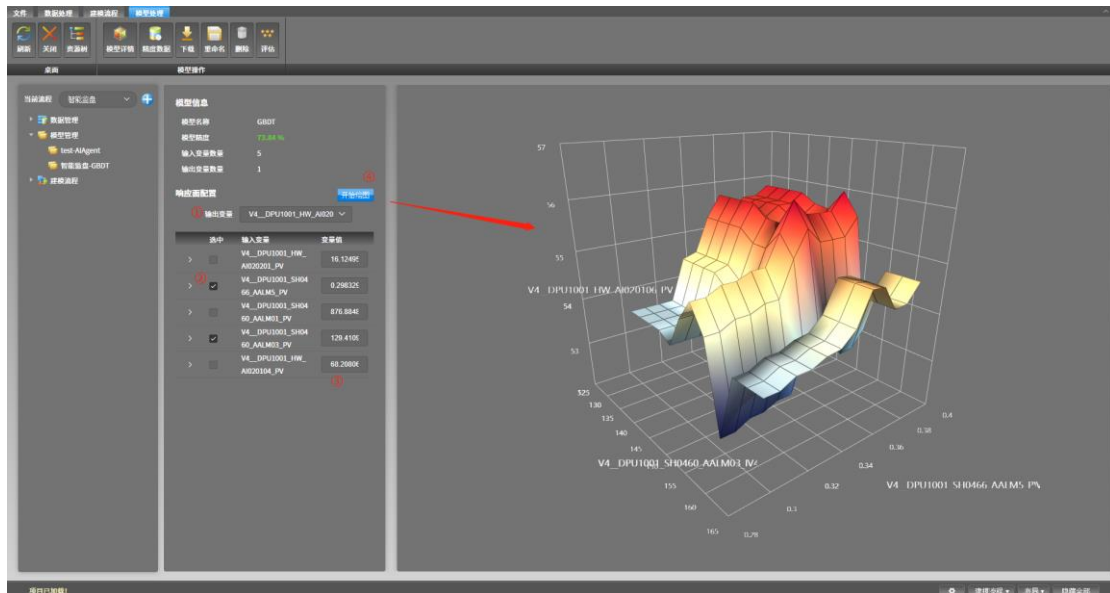


图164 模型响应面信息查看

单击模型处理菜单中的评估按钮，可基于该模型根据指定的输入变量值，给出输出变量的预测值，如图 165 所示。

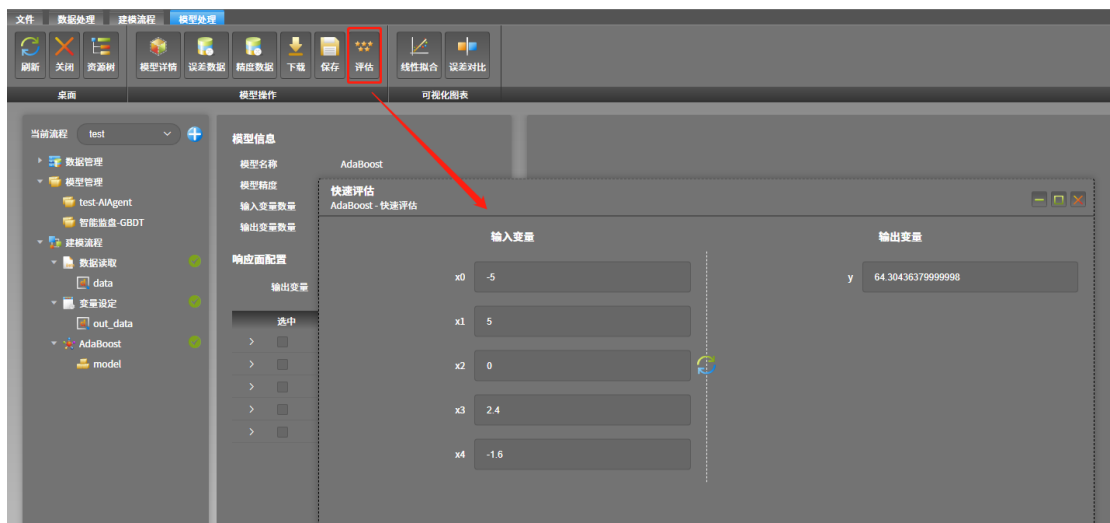


图165 快速评估

对于模型训练节点执行完成得到的模型，还可通过模型菜单中的误差数据查看模型对训练集的预测效果，辅助用户对模型精度是否满足自身需求进行判定。除此之外还支持用户基于该数据绘制包含拟合图、误差对比图在内的可视化图表辅助进行分析，如图 166 所示。



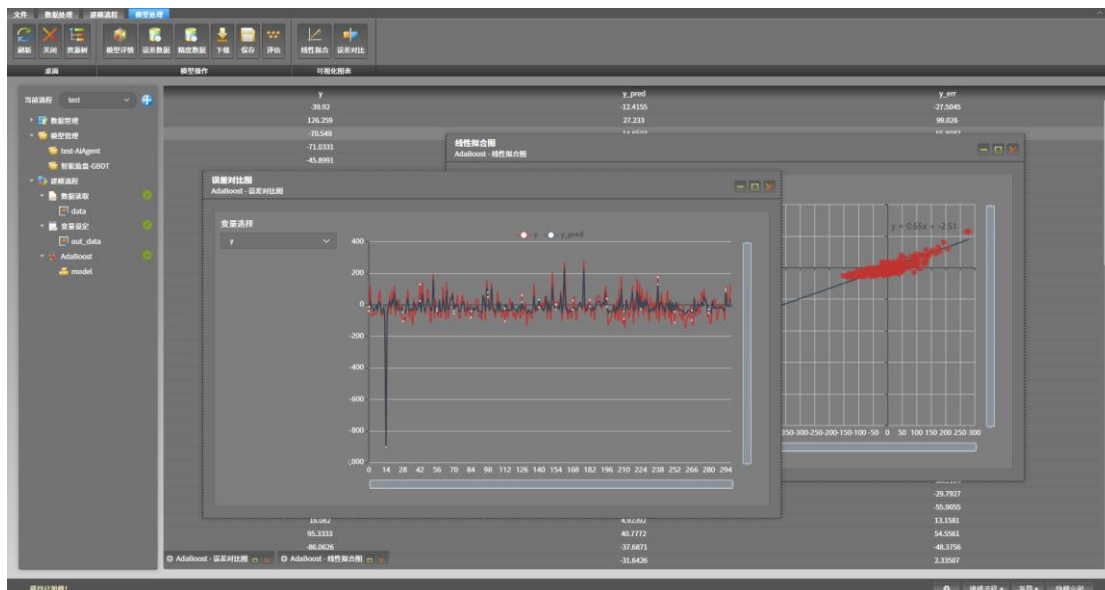


图166 模型训练结果数据查看

5 分析节点

在 DTEmpower V2022R2 版本中，全部分析节点的功能层级统一划分为三级结构。其中，一级功能代表数据建模过程中的不同阶段，共包含 8 个节点，分别为数据载入、数据操作、数据清理、特征生成、敏感性分析、回归训练、智能监盘和模型操作。每个一级功能内包含 1 个或多个二级功能，二级功能之间的关系基本保持独立。三级功能是可独立实现功能的最小单位，其可被拖拽至画布中用于搭建建模流程。一级、二级、三级功能在工具箱中的层级划分如图 167 所示。





图167 分析节点功能层级划分

5.1 数据载入

5.1.1 数据读取

数据读取节点可以从资源树-数据管理中加载数据集，如图 168 所示。



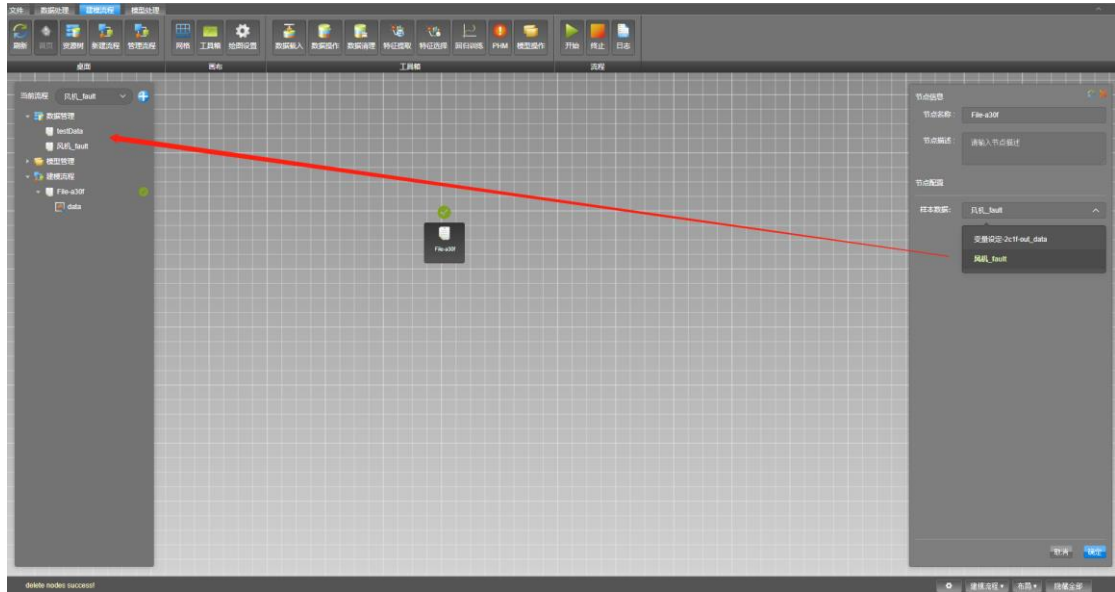


图168 数据读取

5.1.1.1 输出量

数据读取节点的输出量如表 2 所示。

表2 数据读取节点输出量列表

名称	类型	描述
data	DTTable	从资源树-数据管理中读取的数据内容

5.1.1.2 配置参数



图169 数据读取节点属性

数据读取节点配置参数如图 169 所示，各参数描述如下：

- 节点名称：节点在建模流程图中的名称标注；
- 节点描述：节点描述信息，用于备注节点所实现的功能；



- 样本数据：下拉选择资源树-数据管理中的数据集市。

5.1.1.3 使用说明

在数据建模流程中，数据读取节点一般作为起始节点，用于读取原始数据，其下游通常连接接收 dttable 数据作为输入值的功能节点，如变量设定、数据分割、空值处理等。

5.1.2 数据采样

数据采样节点用于快速生成用户指定规则、范围和样本大小的数据集。

5.1.2.1 输出量

数据采样节点的输出量如表 3 所示。

表3 数据采样节点输出量列表

名称	类型	描述
data	DTTable	根据用户指定规则、范围和样本大小生成的数据集

5.1.2.2 配置参数



图170 数据采样节点属性

数据采样节点配置参数如图 170 所示，各参数描述如下：

- 变量名称：用户指定新生成变量的名称；
- 采样下限：用户指定新生成变量范围的下限；
- 采样上限：用户指定新生成变量范围的上限；



- **选择算法：**用户指定数据采样的规则，其中共包括 8 种不同的算法，分别为：FAST(傅里叶振幅灵敏度)、finite_diff(基于导数的全局灵敏度)、fractional_factorial(正交因子)、gaussian(均匀高斯)、LHS(拉丁超立方)、saltelli(sobol 序列的 Saltelli 扩展)、sobol(sobol 序列)、uniform(均匀随机采样)等。
- **采样数量：**用户指定生成数据集的样本大小，应注意对应生成规则的不同，采样数量不等同于生成数据集的行数；
- **随机种子：**可以为空或输入正整数，后者用于指定算法初始值生成的规则。

5.1.2.3 使用说明

在数据建模流程中，数据采样节点一般作为起始节点，用于生成用户指定规则、范围和样本大小的数据集。其下游通常连接接收 dtable 数据作为输入值的功能节点，如变量设定、数据分割、空值处理等。

5.2 数据设定

5.2.1 变量设定

变量设定节点用于指定模型训练中的变量类型，分为输入变量、输出变量和其他变量三类。其中输入变量为最终训练模型的输入变量，输出变量为最终训练模型的输出变量，其他变量为时间、索引等不参与模型训练的变量。变量设定节点可以视为对数据集的“列设定”。

5.2.1.1 输入量

变量设定节点输入量如表 4 所示。

表4 变量设定节点输入量列表

名称	类型	描述
in_data	DTable	需要指定或变更变量类型的数据

5.2.1.2 输出量

变量设定节点输出量如表 5 所示。

表5 变量设定节点输出量列表



名称	类型	描述
out_data	DTable	指定变量类型的数据

5.2.1.3 配置参数



图171 变量设定节点属性

变量设定节点配置参数如图 171 所示，各参数描述如下：

- 节点名称：节点在建模流程图中的名称标注；
- 节点描述：节点描述信息，用于备注节点所实现的功能；
- 变量类型列表：各类型变量名称。

需要注意的是，当该节点未执行时，变量名称对于该节点是未知的。当该节点被执行时，系统会自动弹出变量类型列表，用户在弹窗中实现变量列的设定操作，如图 172 所示。变量列表选择中支持 Shift 快捷键批量选择功能。





图172 变量类型列表

5.2.1.4 使用说明

在数据建模流程中，变量设定节点一般位于输出量为 DTable 节点下游，用于指定变量类型，如数据载入节点。其下游通常连接接收 DTable 数据作为输入值的功能节点，如数据清洗、模型训练、特征生成等。

5.2.2 数据分割

数据分割节点能够将原始数据集按指定比例分割为两部分，可分别用作模型训练和模型测试。数据分割节点可以视为对数据集的“行设定”。

5.2.2.1 输入量

数据分割节点输入量如表 6 所示。

表6 数据分割节点输入量列表

名称	类型	描述
in_data	DTable	需要进行分割的数据集

5.2.2.2 输出量

数据分割节点输出量如表 7 所示。

表7 数据分割节点输出量列表

名称	类型	描述
train_data	DTable	分割后的训练数据集
test_data	DTable	分割后的测试数据集



5.2.2.3 配置参数



图173 数据分割节点属性

数据分割节点配置参数如图 173 所示，各参数描述如下：

- 节点名称：节点在建模流程图中的名称标注；
- 节点描述：节点描述信息，用于备注节点所实现的功能；
- 测试集占比：分割后的测试集算例占总算例的比例，为 0~1 之间的数值；
- 随机种子：进行数据分割时的随机种子。

5.2.2.4 使用说明

在数据建模流程中，数据分割节点一般位于输出量为 DTable 节点下游，用于变量筛选，如数据载入节点。其下游通常连接接收 DTable 数据作为输入值的功能节点，如模型训练、模型对比等。

5.3 数据筛选

5.3.1 变量范围

变量范围节点支持用户对每个变量的上下限进行设定，执行变量范围节点后，数据集中不满足用户指定范围的算例将被移除。

5.3.1.1 输入量

变量范围节点输入量如表 8 所示。

表8 变量范围节点输入量列表



名称	类型	描述
in_data	DTTable	原始范围数据

5.3.1.2 输出量

变量范围节点输出量如表 9 所示。

表9 变量范围节点输出量列表

名称	类型	描述
out_data	DTTable	移除不满足设定范围算例的数据

5.3.1.3 配置参数



图174 变量范围节点属性

变量范围节点配置参数如图 174 所示，各参数描述如下：

- 节点名称：节点在建模流程图中的名称标注；
- 节点描述：节点描述信息，用于备注节点所实现的功能；
- 变量范围列表：各变量上下限列表。

需要注意的是，当该节点未执行时，原始变量范围对于该节点是未知的。当该节点被执行时，系统会自动弹出变量范围列表，且默认填入参数当前上下限。用户在弹窗中实现变量范围的设定操作，如图 175 所示。





图175 变量范围列表

5.3.1.4 使用说明

在数据建模流程中，变量范围节点一般位于输出量为 DTable 节点下游，用于数据筛选，其下游通常连接接收 DTable 数据作为输入值的功能节点，如变量设定、数据清洗、模型训练等。

5.3.2 变量筛选

在变量筛选节点中，用户可指定数据中被保留的变量。

5.3.2.1 输入量

变量筛选节点输入量如表 10 所示。

表10 变量筛选节点输入量列表

名称	类型	描述
in_data	DTable	需要保留某些变量的数据

5.3.2.2 输出量

变量筛选节点输出量如表 11 所示。

表11 变量筛选节点输出量列表

名称	类型	描述
out_data	DTable	保留指定变量的数据



5.3.2.3 配置参数

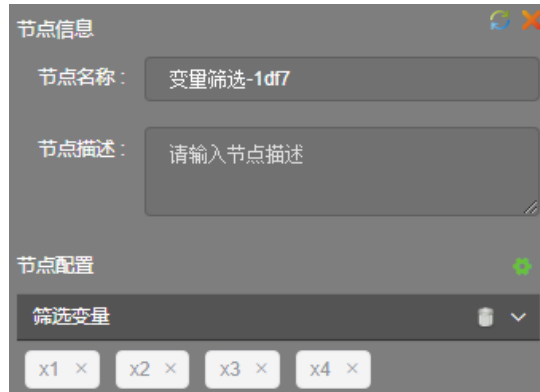


图176 变量筛选节点属性

变量筛选节点配置参数如图 176 所示，各参数描述如下：

- 节点名称：节点在建模流程图中的名称标注；
- 节点描述：节点描述信息，用于备注节点所实现的功能；
- 筛选变量：需要被保留的变量。

需要注意的是，当该节点未执行时，变量名称对于该节点是未知的。当该节点被执行时，系统会自动弹出筛选变量列表，用户在弹窗中实现变量筛选操作，如图 177 所示。



图177 筛选变量列表



5.3.2.4 使用说明

在数据建模流程中，变量筛选节点一般位于输出量为 DTable 节点下游，用于变量筛选，如数据载入节点。其下游通常连接接收 DTable 数据作为输入值的功能节点，如数据清洗、模型训练、特征生成等。

5.3.3 变量剔除

在变量剔除节点中，用户可指定数据中被移除的变量。

5.3.3.1 输入量

变量剔除节点输入量如表 12 所示。

表12 变量剔除节点输入量列表

名称	类型	描述
in_data	DTable	需要删除某些变量的数据

5.3.3.2 输出量

变量剔除节点输出量如表 13 所示。

表13 变量剔除节点输出量列表

名称	类型	描述
out_data	DTable	删除指定变量的数据

5.3.3.3 配置参数



图178 变量剔除节点属性

变量剔除节点配置参数如图 178 所示，各参数描述如下：

- 节点名称：节点在建模流程图中的名称标注；
- 节点描述：节点描述信息，用于备注节点所实现的功能；



- 剔除变量列表：需要被移除的变量。

需要注意的是，当该节点未执行时，变量名称对于该节点是未知的。当该节点被执行时，系统会自动弹出剔除变量列表，用户在弹窗中实现变量剔除操作，如图 179 所示。

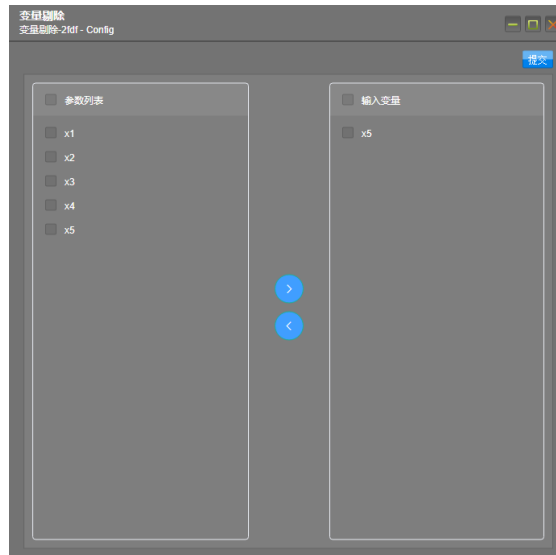


图179 剔除变量列表

5.3.3.4 使用说明

在数据建模流程中，变量剔除节点一般位于输出量为 DTable 节点下游，用于变量筛选，如数据载入节点。其下游通常连接接收 DTable 数据作为输入值的功能节点，如数据清洗、模型训练、特征生成等。

5.3.4 稀释采样

稀释采样节点根据用户指定的采样规则对输入数据进行采样，以实现降低数据规模的效果。

5.3.4.1 输入量

稀释采样节点输入量如表 14 所示。

表14 稀释采样节点输入量列表

名称	类型	描述
in_data	DTable	需要降低规模的数据



5.3.4.2 输出量

稀释采样节点输出量如表 15 所示。

表15 稀释采样节点输出量列表

名称	类型	描述
in_data	DTTable	降低规模后的数据

5.3.4.3 配置参数

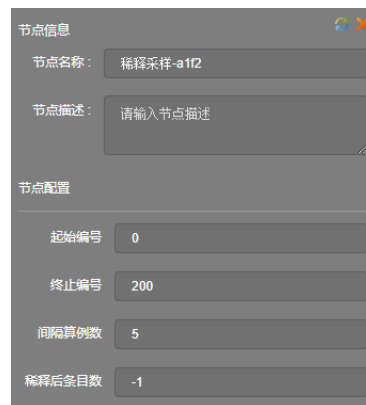


图180 稀释采样节点属性

稀释采样节点配置参数如图 180 所示，各参数描述如下：

- 节点名称：节点在建模流程图中的名称标注；
- 节点描述：节点描述信息，用于备注节点所实现的功能；
- 起始编号：起始采样算例编号；
- 终止编号：终止采样算例编号；
- 间隔算例数：稀释采样时的采样间隔，即每隔多少个算例进行一次采样；
- 稀释后条目数：稀释后的数据规模，当指定为-1 时，则完全按照起始编号、终止编号、间隔算例数进行稀释采样。

5.3.4.4 使用说明

在数据建模流程中，稀释采样节点一般位于输出量为 DTTable 节点下游，用于数据稀释，如数据载入节点。其下游通常连接接收 DTTable 数据作为输入值的功能节点，如模型训练、模型对比等。



5.3.5 空值处理

空值处理节点根据用户指定的处理方式对包含空值的算例进行操作，以实现空值算例的删除等功能。

5.3.5.1 输入量

空值处理节点输入量如表 16 所示。

表16 空值处理节点输入量列表

名称	类型	描述
in_data	DTTable	包含空值的数据

5.3.5.2 输出量

空值处理节点输出量如表 17 所示。

表17 空值处理节点输出量列表

名称	类型	描述
out_data	DTTable	不包含空值的数据

5.3.5.3 配置参数

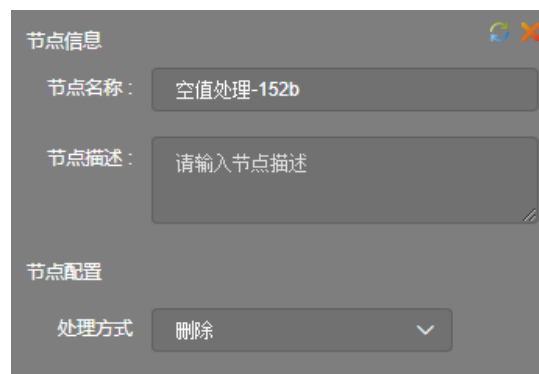


图181 空值处理节点属性

空值处理节点配置参数如图 181 所示，各参数描述如下：

- 节点名称：节点在建模流程图中的名称标注；
- 节点描述：节点描述信息，用于备注节点所实现的功能；
- 处理方式：对包含空值算例的处理方法。



5.3.5.4 使用说明

在数据建模流程中，空值处理节点一般位于输出量为 DTable 节点下游，用于样本筛选，如数据载入节点。其下游通常连接接收 DTable 数据作为输入值的功能节点，如变量设定、数据清洗、特征工程等。

5.3.6 自定义剔除

自定义剔除节点根据用户指定的规则，移除数据中不满足规则的算例。

5.3.6.1 输入量

自定义剔除节点输入量如表 18 所示。

表18 自定义剔除节点输入量列表

名称	类型	描述
in_data	DTable	需要移除满足特定规则算例的数据

5.3.6.2 输出量

自定义剔除节点输出量如表 19 所示。

表19 自定义剔除节点输出量列表

名称	类型	描述
out_data	DTable	移除满足特定规则算例的数据

5.3.6.3 配置参数



图182 自定义剔除节点属性

自定义剔除节点配置参数如图 182 所示，各参数描述如下：

- 节点名称：节点在建模流程图中的名称标注；
- 节点描述：节点描述信息，用于备注节点所实现的功能；



- 筛选表达式：依托于变量组成的筛选规则计算公式。

5.3.6.4 使用说明

在数据建模流程中，自定义剔除节点一般位于输出量为 DTable 节点下游，用于算例筛选，如数据载入节点。其下游通常连接接收 DTable 数据作为输入值的功能节点，如模型训练、变量设定等。

5.4 表格处理

5.4.1 重命名

默认情况下，系统会自动将表头字符串作为变量名称，重名节点用于对变量名称进行变更。

5.4.1.1 输入量

重命名节点输入量如表 20 所示。

表20 重命名节点输入量列表

名称	类型	描述
in_data	DTable	带有原始变量名称的数据

5.4.1.2 输出量

重命名节点输出量如表 21 所示。

表21 重命名节点输出量列表

名称	类型	描述
out_data	DTable	变更变量名称后的数据



5.4.1.3 配置参数



节点信息

节点名称: 重命名

节点描述: 请输入节点描述

节点配置

变量名称	重命名 !	注释
Time	Time	
DCS2_20HFE20 AN001CE	c1	
DCS2_20HFE10 AN001CE	c2	

图183 重命名节点属性

重命名节点配置参数如图 183 所示，各参数描述如下：

- 节点名称：节点在建模流程图中的名称标注；
- 节点描述：节点描述信息，用于备注节点所实现的功能；
- 变量名称列表：原变量名称和重命名后的变量名称。

需要注意的是，当该节点未执行时，原变量名对于该节点是未知的。当该节点被执行时，系统会自动弹出变量名称列表，用户在弹窗中实现变量的重命名操作，如图 184 所示。



重命名
重命名 - Config

提交

变量名称	重命名 !	注释	操作
Time	Time		上移 下移
DCS2_20HFE20 AN001CE	c1		上移 下移
DCS2_20HFE10 AN001CE	c2		上移 下移

图184 重命名变量名称列表



5.4.1.4 使用说明

在数据建模流程中，重命名节点一般位于输出量为 DTable 节点下游，用于改变变量名称，其下游通常连接接收 DTable 数据作为输入值的功能节点，如变量设定、数据清洗、模型训练等。

5.4.2 数据存储

数据存储用于保存流程运行过程中的 data 计算结果，并提供了保存至资源树或保存至本地等多种存储方式。

5.4.2.1 输入量

数据存储节点输入量如表 22 所示。

表22 数据存储节点输入量列表

名称	类型	描述
in_data	DTable	需要保存的数据结果

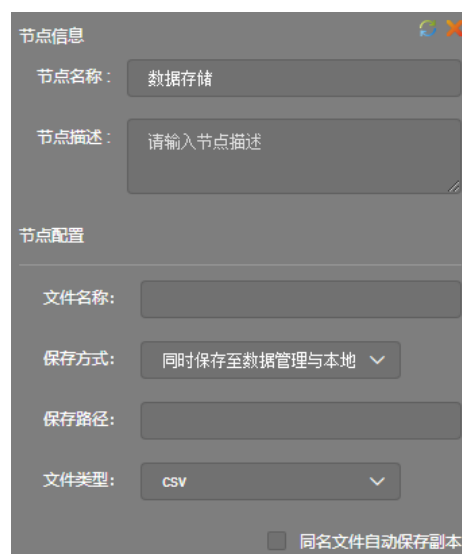
5.4.2.2 输出量

数据存储节点输出量如表 23 所示。

表23 数据存储节点输出量列表

名称	类型	描述
out_data	DTable	已保存的数据结果

5.4.2.3 配置参数



节点信息

节点名称: 数据存储

节点描述: 请输入节点描述

节点配置

文件名称:

保存方式: 同时保存至数据管理与本地

保存路径:

文件类型: CSV

同名文件自动保存副本



图185 数据存储节点属性

数据存储节点配置参数如图 185 所示，各参数描述如下：

- 节点名称：节点在建模流程图中的名称标注；
- 节点描述：节点描述信息，用于备注节点所实现的功能；
- 文件名称：保存数据结果的名称；
- 保存方式：可选择保存至左侧资源树的数据管理中，或选择保存至本地，也可以选择二者同时保存；
- 保存路径：当选择保存至本地时，需要指定保存绝对路径；
- 文件类型：当选择保存至本地时，需要指定保存数据文件的类型，可选的类型有 csv、xlsx 以及 DTable。

5.4.2.4 使用说明

在数据建模流程中，数据存储节点一般位于输出量为 DTable 节点下游，用于保存计算结果，且后续不再连接其他节点。

5.5 时序数据操作

5.5.1 时序变量设定

时序变量设定节点用于指定时间序列模型训练中的变量类型，分为时间变量、预测变量和外源变量三类。其中时间变量为时序模型的对应时间索引列，预测变量为时序模型用于预测的变量，外源变量为影响预测变量的其他变量。

时序变量设定节点可视为对时间序列数据集的“列设定”，且与用于回归训练的变量设定节点有一定的相似性。

5.5.1.1 输入量

时序变量设定节点输入量如表 24 所示。

表24 时序变量设定节点输入量列表

名称	类型	描述
in_data	DTable	需要指定或变更变量类型的数据

5.5.1.2 输出量

时序变量设定节点输出量如表 25 所示。



表25 时序变量设定节点输出量列表

名称	类型	描述
out_data	TS-DTTable	指定变量类型的时间序列数据

5.5.1.3 配置参数



图186 时序变量设定节点属性

时序变量设定节点配置参数如图 186 所示，各参数描述如下：

- 节点名称：节点在建模流程图中的名称标注；
- 节点描述：节点描述信息，用于备注节点所实现的功能；
- 变量类型列表：各类型变量名称，包括时间变量、预测变量与外源变量三类；
- 时序数据描述：展示时序数据的基本信息，如起始时间、结束时间、样本个数，以及是否满足校验规则等。

需要注意的是，当该节点未执行时，变量名称对于该节点是未知的。当该节点被执行时，系统会自动弹出变量类型列表，用户在弹窗中实现变量列的设定操作，如图 187 所示。时序变量列表选择中支持 Shift 快捷键批量选择功能。





图187 时序变量设定配置

5.5.1.4 使用说明

在数据建模流程中，时序变量设定节点一般位于输出量为 DTable 节点下游，用于指定变量类型，如数据载入节点。其下游通常连接接收时间序列数据作为输入值的功能节点，如时序预处理、时序训练算法等。

进行过时序变量设定的数据具有独特的结构形式：一般的数据集用递增的序号标记每个样本，而时序数据集用日期/时间标记每个样本，如图 188 所示。

	A
	123 y
2019-01-06 07:05:00	945.7464
2019-01-06 07:06:00	946.1856
2019-01-06 07:07:00	946.7982
2019-01-06 07:08:00	946.9187
2019-01-06 07:09:00	947.3391
2019-01-06 07:10:00	947.6313

图188 时序数据集用日期/时间标记每个样本

另外需要注意的是，变量类型列表中，各变量设定彼此之间不能重复。此外还存在一定的数量和类型限制，具体如下：

- 时间变量：有且仅有一个变量，且支持的变量类型为日期型（时间戳）或序列型（等间隔的数值）；



- 预测变量：有且仅有一个变量，且支持的变量类型为数值型；
- 外源变量：可以有 0 个、1 个或多个外源变量，且支持的变量类型为数值型。

5.5.2 时序预处理

时序模型训练算法对于数据集有比较严格的限制，如连续采样且时间间隔均匀等。原始数据集往往无法满足此类算法的需要，因此必须通过时序预处理节点将其完善为满足规则的时序数据集。

5.5.2.1 输入量

时序预处理节点输入量如表 26 所示。

表26 时序预处理节点输入量列表

名称	类型	描述
in_data	TS-DTTable	需要进行预处理的时序数据

5.5.2.2 输出量

时序预处理节点输出量如表 27 所示。

表27 时序预处理节点输出量列表

名称	类型	描述
out_data	TS-DTTable	符合校验规则的时间序列数据



5.5.2.3 配置参数




图189 时序预处理节点属性

时序变量设定节点配置参数如图 189 所示，各参数描述如下：

- 节点名称：节点在建模流程图中的名称标注；
- 节点描述：节点描述信息，用于备注节点所实现的功能；
- 起始时间：显示时序数据的起始时间，即第一个样本对应的时间变量；
- 结束时间：显示时序数据的结束时间，即最后一个样本对应的时间变量；
- 样本个数：显示数据集行数；
- 通过校验：显示预处理后的数据集是否满足校验规则。

需要注意的是，当该节点未执行时，数据信息对于该节点是未知的。当该节点被执行时，系统会自动弹出配置面板，用户在弹窗中实现时序预处理的设定操作，如图 190 所示。



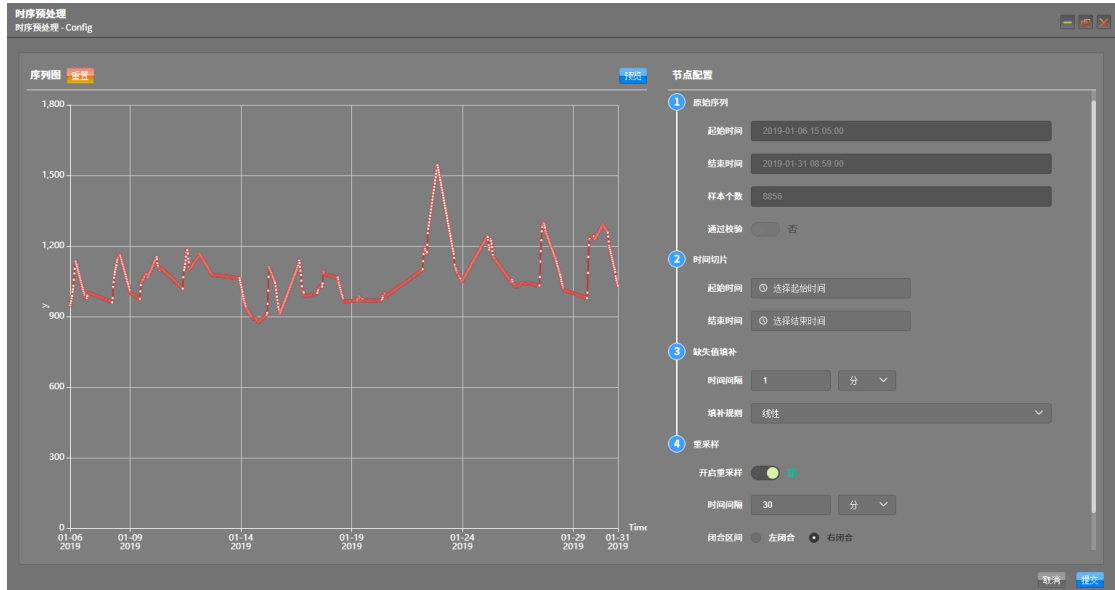


图190 时序预处理配置

在图 190 的配置弹窗中，整体可分为 2 大部分：即左侧的序列图，以及右侧的节点配置面板。

在序列图的顶部，提供“预览”以及“重置”2 个功能按钮。其中“预览”按钮会通过右侧节点配置面板对数据集进行处理，处理后的结果将显示于左侧序列图。而重置按钮则会忽略处理后的数据，直接绘制原始数据集对应的序列图。默认状态下，序列图显示原始数据集的结果。

在序列图上，鼠标悬浮会显示当前时刻及其对应的预测变量值。单击鼠标右键可唤出快速时间切片菜单，并提供“设为起点”和“设为终点”2 个选项。当起点和终点同时完成设定后，会自动标出高亮区域显示截取的时间切片位置，如图 191 所示。此时点击“预览”按钮即可显示切片后的序列图。



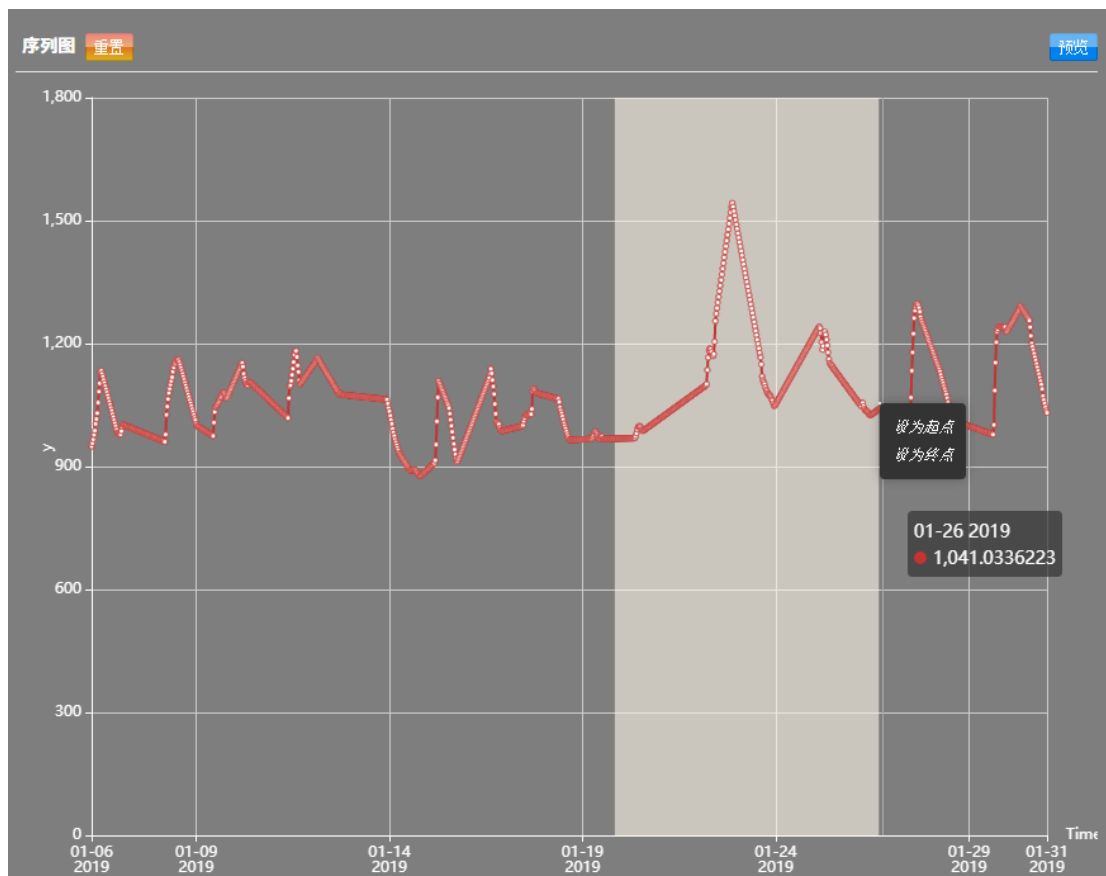


图191 序列图快速切片

节点配置面板分为4个环节：原始序列、时间切片、缺失值填补以及重采样，各环节之间存在先后顺序关系。

对于原始序列，其显示了上游节点传入的时序数据基本信息，包括起始时间、结束时间、样本个数，以及是否通过校验。此处的信息只用于展示而不可修改，与左侧初始的序列图或重置后的序列图相对应。

对于时间切片，可用于设置起始时间和结束时间。以起始时间设置为例，提供了三种方式进行修改：①手动输入形如“yyyy-MM-dd HH:mm:ss”的时间形式；②通过时间选择器，选择某个时刻；③在序列图中，通过右键菜单设定。结束时间的设置与起始时间相似。

对于缺失值填补，需要同时设定时间间隔，以及当对应时刻不存在原始数据时的填补规则。其中，时间间隔应与传感器的采样频率相匹配。填补规则提供了如前值（与前一个有效值一致）、后值（与后一个有效值一致）、固定值（用户指



定的有效值)、平均值(全局有效值的平均数)、线性(前后相邻有效值的平均数)等 5 种方式。进行过缺失值填补的时序数据必然会满足校验规则。

对于重采样,是于完成缺失值填补后,需要调整算法训练用的采样频率时所使用。此处的重采样限定为下采样,即新的时间间隔须大于原始时间间隔(缺失值填补中设定值)。同时提供相应的闭合区间与聚合规则配置选项。



图192 缺失值填补与重采样

5.5.2.4 使用说明

在时序建模流程中,时序预处理节点一般位于输出量为时序数据的节点下游,用于时序数据的起止时间设定、缺失值填补、重采样等,如时序变量设定节点。其下游通常连接接收时序数据作为输入值的功能节点,如时序训练算法、时序模型更新等。

5.6 数据清理

5.6.1 智能清理算法 AIOD

DTEmpower 提供了智能数据清理算法 AIOD,该算法将 20 多种常见算法和自研调度算法有效结合。具有强大的默认设置和自适应性,支持用户一键启动数据清理流程,极大程度降低用户使用门槛的情况下,满足了绝大多数异常点识别场景。

5.6.1.1 输入量

AIOD 节点输入量如表 28 所示。



表28 AIOD 节点输入量列表

名称	类型	描述
in_data	DTTable	需要进行数据清洗的原始数据

5.6.1.2 输出量

AIOD 节点输出量如表 29 所示。

表29 AIOD 节点输出量列表

名称	类型	描述
normal_data	DTTable	非异常点算例组成的数据集
outlier_data	DTTable	异常点算例组成的数据集

5.6.1.3 配置参数



节点信息

节点名称: AIOD

节点描述: 请输入节点描述

节点配置

异常点检测等级: 1

异常点数量: -1

图193 AIOD 节点属性

AIOD 节点配置参数如图 193 所示，各参数描述如下：

- 节点名称：节点在建模流程图中的名称标注；
- 节点描述：节点描述信息，用于备注节点所实现的功能；
- 异常点检测等级：分为 1、2、3 共三个检测等级，等级越高，对异常点的识别就越敏感，相应的异常风险评分也更高；
- 异常点数量：可设置为-1（默认）或正整数。当设置为正整数时，节点运行后会从最高的异常风险评分开始，依次向下筛选对应数量的样本并



剔除（即认定为异常点）；当设置为-1 时，节点运行后会弹出可视化配置弹窗，用户可手动筛选异常样本并剔除。

需要注意的是，当该节点未执行时，各算例情况对于该节点是未知的。当该节点被执行且异常点数量设置为-1 时，系统会自动弹出异常风险等级列表，用户在弹窗中对异常点检测结果进行人为干预，如图 194 所示。

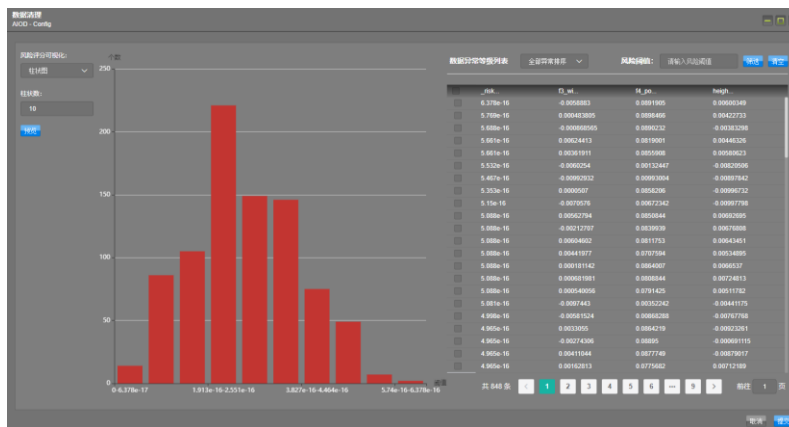


图194 AIOD 异常风险等级列表

如图 194 所示，异常风险等级列表右侧为节点输入数据所有算例异常风险评分从高到低排序后的算例列表，其中 `_risk_scores_` 为算例对应的异常风险评分。左侧为 2 维数据可视化图表，用以辅助用户进行异常算例的选择。右侧列表中被选中的算例被视为异常算例，节点执行完成后，被存储至输出量 `outlier_data` 数据集之中。

AIOD 中提供了 3 种不同的风险评分可视化形式，分别为柱状图、箱线图以及散点图。当右侧异常点选中发生改变时，通过点击左侧的“预览”按钮可以在左侧图形中同步显示。

- 柱状图：默认的可视化形式，将风险评分数值按照从低到高的形式进行排列并分段，并分别计算各段的频率；分割的柱状图可由用户指定；
- 箱线图：将风险评分用箱线图的形式显示；选中的异常点将在图形中被剔除；
- 散点图：默认显示风险评分与样本序号之间的二维散点图，也支持用户切换 X 轴与 Y 轴的显示变量；选中的异常点将在图形中高亮显示；用



用户可双击某一算例将其指定为异常算例，也可双击被指定为异常的算例，取消异常指定；支持通过图形右上角的框选按钮进行批量框选。

在弹窗右上角，还支持基于风险阈值的批量选择，将选中全部风险阈值大于填入数值的样本为异常点。清空按钮则会取消全部异常点选中。

5.6.1.4 使用说明

在数据建模流程中，AIOD 节点一般位于输出量为 DTable、且完成了变量类型指定的节点下游，与接收 DTable 作为训练数据的模型训练节点之间，用以移除数据中的异常算例，提升模型训练效果。

5.6.2 基元算法清理

DTEmpower 中还提供有 20 种基元算法清理节点，方便专业用户根据数据集特点手动选择最优的清理算法。数据清理基元算法名称如表 30 所示。

表30 数据清理基元算法

CBLOF	Forest	LODA	ProximityABOD
COF	HBOS	LOF	REOD_ICA
EHPD	KNN	LSCPOE	REOD_PCA
Elliptic	LMDD	MCD	SOD
EnseFB	LOCI	OCSVM	SOS

5.6.2.1 输入量

基元算法清理节点输入量如表 31 所示。

表31 基元算法清理节点输入量列表

名称	类型	描述
in_data	DTable	需要进行数据清洗的原始数据

5.6.2.2 输出量

基元算法清理节点输出量如表 32 所示。

表32 基元算法清理节点输出量列表

名称	类型	描述
normal_data	DTable	非异常点算例组成的数据集



5.6.2.4 使用说明

在数据建模流程中，基元算法清理节点一般位于输出量为 DTable、且完成了变量类型指定的节点下游，与接收 DTable 作为训练数据的模型训练节点之间，用以移除数据中的异常算例，提升模型训练效果。选择不同的基元算法会计算出不同的异常风险评分(_risk_scores_)，其余操作方式则完全一致。

5.7 数据聚类

5.7.1 KMeans

KMeans 节点数据聚类功能，可以通过无监督学习的方式将一堆没有标签的数据自动划分成几类，且保证同一类的数据有相似的特征。

5.7.1.1 输入量

KMeans 节点输入量如表 33 所示。

表33 KMeans 节点输入量列表

名称	类型	描述
in_data	DTable	需进行聚类分析的数据

5.7.1.2 输出量

KMeans 节点输出量如表 34 所示。

表34 KMeans 节点输出量列表

名称	类型	描述
out_data	DTable	在原始数据基础上新增聚类变量
model	Model	聚类后得到的模型



5.7.1.3 配置参数

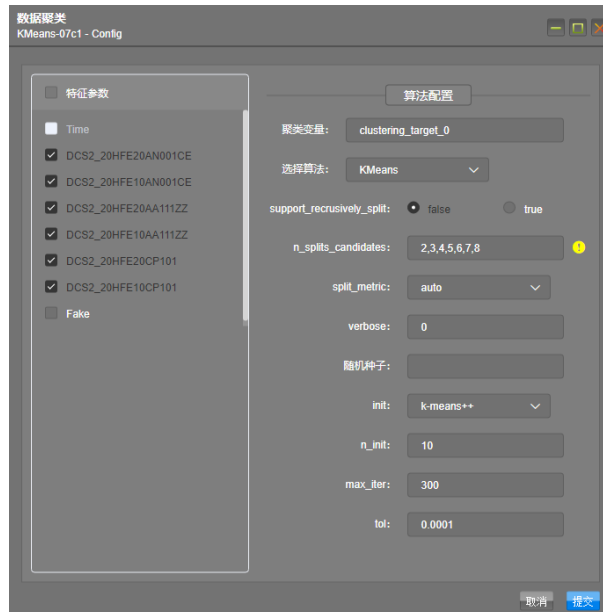


图197 KMeans 节点属性

KMeans 节点配置参数如图 197 所示，各参数描述如下：

- 节点名称：节点在建模流程图中的名称标注；
- 节点描述：节点描述信息，用于备注节点所实现的功能；
- 特征参数列表：勾选数据集内用于聚类分析的特征参数；
- 聚类变量：聚类分析后新生成的变量列名；
- 选择算法：KMeans；
- support_recursively_split：定义当前方法是否支持递归拆分；
- n_splits_candidates：指定生成聚合类数量的候选值；
- split_metric：指定在轮廓分数中定义无监督聚类度量的方式；
- n_init：使用不同质心种子运行 k-means 算法的时间；
- max_iter：单次运行的 k-means 算法的最大迭代次数；
- tol：聚类中心连续迭代收敛判断。

5.7.1.4 使用说明

在数据建模流程中，KMeans 节点一般位于输出量为 DTable 的节点下游，模型（知识）聚合节点的上游，用于将数据根据内在性质分成不同的聚合类，进而针对特定聚合类作进一步分析。



5.7.2 Batch-KMeans

Batch-KMeans 节点数据聚类功能，可以通过无监督学习的方式将一堆没有标签的数据自动划分成几类，且保证同一类的数据有相似的特征。

5.7.2.1 输入量

Batch-KMeans 节点输入量如表 35 所示。

表35 Batch-KMeans 节点输入量列表

名称	类型	描述
in_data	DTTable	需进行聚类分析的数据

5.7.2.2 输出量

Batch-KMeans 节点输出量如表 36 所示。

表36 Batch-KMeans 节点输出量列表

名称	类型	描述
out_data	DTTable	在原始数据基础上新增聚类变量
model	Model	聚类后得到的模型

5.7.2.3 配置参数

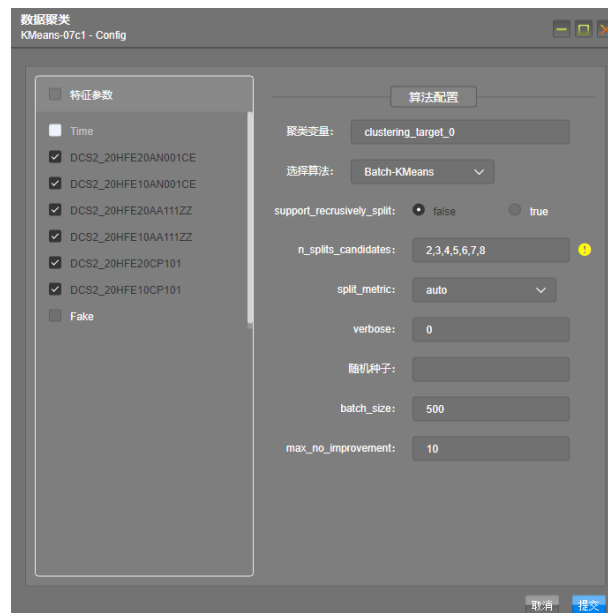


图198 Batch-KMeans 节点属性

Batch-KMeans 节点配置参数如图 198 所示，各参数描述如下：



- 节点名称：节点在建模流程图中的名称标注；
- 节点描述：节点描述信息，用于备注节点所实现的功能；
- 特征参数列表：勾选数据集内用于聚类分析的特征参数；
- 聚类变量：聚类分析后新生成的变量列名；
- 选择算法：Batch-KMeans；
- support_recursively_split：定义当前方法是否支持递归拆分；
- n_splits_candidates：指定生成聚合类数量的候选值；
- split_metric：指定在轮廓分数中定义无监督聚类度量的方式；
- batch_size：指定小批量的大小；
- max_no_improvement：控制提前终止迭代。

5.7.2.4 使用说明

在数据建模流程中，Batch-KMeans 节点一般位于输出量为 DTable 的节点下游，模型（知识）聚合节点的上游，用于将数据根据内在性质分成不同的聚合类，进而针对特定聚合类作进一步分析。

5.7.3 dr_gmm

dr_gmm 节点数据聚类功能，可以通过无监督学习的方式将一堆没有标签的数据自动划分成几类，且保证同一类的数据有相似的特征。

5.7.3.1 输入量

dr_gmm 节点输入量如表 37 所示。

表37 dr_gmm 节点输入量列表

名称	类型	描述
in_data	DTable	需进行聚类分析的数据

5.7.3.2 输出量

dr_gmm 节点输出量如表 38 所示。

表38 dr_gmm 节点输出量列表

名称	类型	描述
out_data	DTable	在原始数据基础上新增聚类变量



5.7.3.3 配置参数

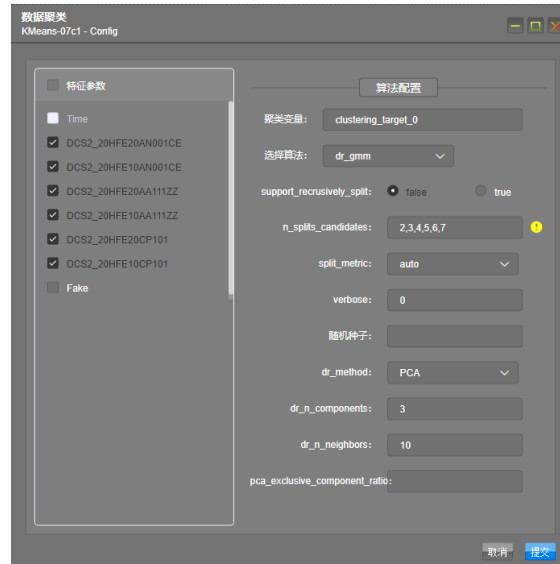


图199 dr_gmm 节点属性

dr_gmm 节点配置参数如图 199 所示，各参数描述如下：

- 节点名称：节点在建模流程图中的名称标注；
- 节点描述：节点描述信息，用于备注节点所实现的功能；
- 特征参数列表：勾选数据集内用于聚类分析的特征参数；
- 聚类变量：聚类分析后新生成的变量列名；
- 选择算法：dr_gmm；
- support_recursively_split：定义当前方法是否支持递归拆分；
- n_splits_candidates：指定生成聚合类数量的候选值；
- split_metric：指定在轮廓分数中定义无监督聚类度量的方式；
- dr_n_components：指定降维组分数；
- dr_n_neighbors：指定降维邻居数。

5.7.3.4 使用说明

在数据建模流程中，dr_gmm 节点一般位于输出量为 DTable 的节点下游，模型（知识）聚合节点的上游，用于将数据根据内在性质分成不同的聚合类，进而针对特定聚合类作进一步分析。



5.7.4 fcm

fcm 节点数据聚类功能，可以通过无监督学习的方式将一堆没有标签的数据自动划分成几类，且保证同一类的数据有相似的特征。

5.7.4.1 输入量

fcm 节点输入量如表 39 所示。

表39 fcm 节点输入量列表

名称	类型	描述
in_data	DTTable	需进行聚类分析的数据

5.7.4.2 输出量

fcm 节点输出量如表 40 所示。

表40 fcm 节点输出量列表

名称	类型	描述
out_data	DTTable	在原始数据基础上新增聚类变量
model	Model	聚类后得到的模型

5.7.4.3 配置参数



图200 fcm 节点属性

fcm 节点配置参数如图 197 所示，各参数描述如下：

- 节点名称：节点在建模流程图中的名称标注；



- 节点描述：节点描述信息，用于备注节点所实现的功能；
- 特征参数列表：勾选数据集内用于聚类分析的特征参数；
- 聚类变量：聚类分析后新生成的变量列名；
- 选择算法：fcm；
- support_recursively_split：定义当前方法是否支持递归拆分；
- n_splits_candidates：指定生成聚合类数量的候选值；
- split_metric：指定在轮廓分数中定义无监督聚类度量的方式；
- m：在每次迭代中应用于隶属函数的数组求幂
- error：控制提前终止迭代；
- maxiter：控制最大迭代次数；
- judgement：定义群集性能标准。

5.7.4.4 使用说明

在数据建模流程中，fcm 节点一般位于输出量为 DTable 的节点下游，模型（知识）聚合节点的上游，用于将数据根据内在性质分成不同的聚合类，进而针对特定聚合类作进一步分析。

5.7.5 Mean-Shift

Mean-Shift 节点数据聚类功能，可以通过无监督学习的方式将一堆没有标签的数据自动划分成几类，且保证同一类的数据有相似的特征。

5.7.5.1 输入量

Mean-Shift 节点输入量如表 41 所示。

表41 Mean-Shift 节点输入量列表

名称	类型	描述
in_data	DTable	需进行聚类分析的数据

5.7.5.2 输出量

Mean-Shift 节点输出量如表 42 所示。

表42 Mean-Shift 节点输出量列表

名称	类型	描述
----	----	----



out_data	DTTable	在原始数据基础上新增聚类变量
model	Model	聚类后得到的模型

5.7.5.3 配置参数

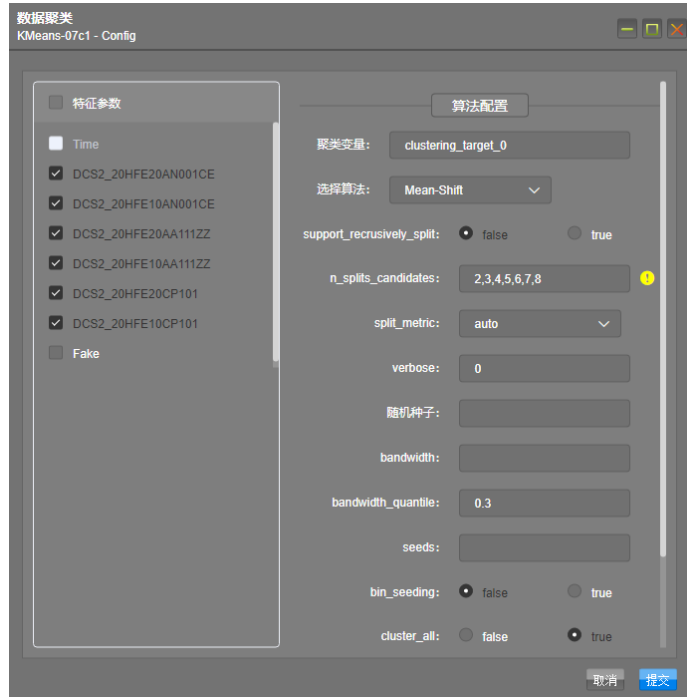


图201 Mean-Shift 节点属性

Mean-Shift 节点配置参数如图 201 所示，各参数描述如下：

- 节点名称：节点在建模流程图中的名称标注；
- 节点描述：节点描述信息，用于备注节点所实现的功能；
- 特征参数列表：勾选数据集内用于聚类分析的特征参数；
- 聚类变量：聚类分析后新生成的变量列名；
- 选择算法：Mean-Shift；
- support_recursively_split：定义当前方法是否支持递归拆分；
- n_splits_candidates：指定生成聚合类数量的候选值；
- split_metric：指定在轮廓分数中定义无监督聚类度量的方式；
- bandwidth_quantile：估计带宽；
- bin_seeding：开启时，减少初始化种子数，加快算法速度；
- cluster_all：开启时，不设置孤立点；



- max_iter: 控制最大迭代次数。

5.7.5.4 使用说明

在数据建模流程中,Mean-Shift 节点一般位于输出量为 DTable 的节点下游,模型(知识)聚合节点的上游,用于将数据根据内在性质分成不同的聚合类,进而针对特定聚合类作进一步分析。

5.7.6 normal_gmm

normal_gmm 节点数据聚类功能,可以通过无监督学习的方式将一堆没有标签的数据自动划分成几类,且保证同一类的数据有相似的特征。

5.7.6.1 输入量

normal_gmm 节点输入量如表 43 所示。

表43 normal_gmm 节点输入量列表

名称	类型	描述
in_data	DTable	需进行聚类分析的数据

5.7.6.2 输出量

normal_gmm 节点输出量如表 44 所示。

表44 normal_gmm 节点输出量列表

名称	类型	描述
out_data	DTable	在原始数据基础上新增聚类变量
model	Model	聚类后得到的模型



5.7.6.3 配置参数

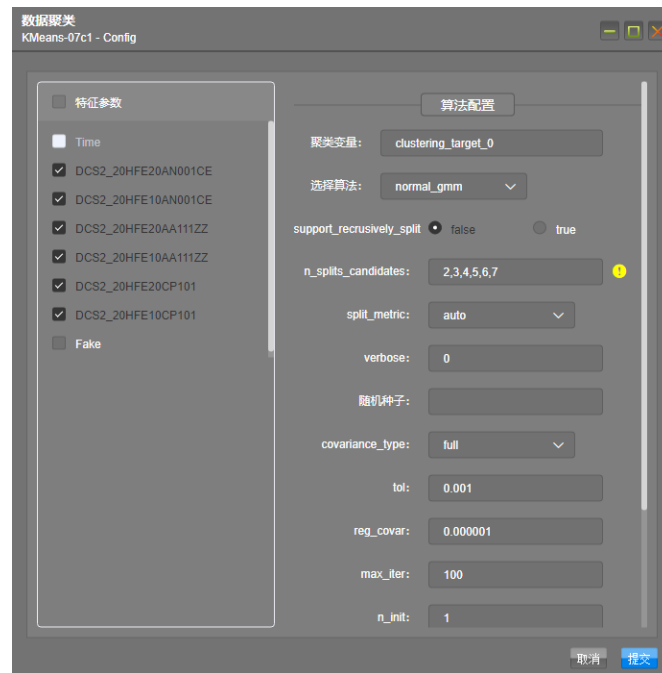


图202 normal_gmm 节点属性

normal_gmm 节点配置参数如图 202 所示，各参数描述如下：

- 节点名称：节点在建模流程图中的名称标注；
- 节点描述：节点描述信息，用于备注节点所实现的功能；
- 特征参数列表：勾选数据集内用于聚类分析的特征参数；
- 聚类变量：聚类分析后新生成的变量列名；
- 选择算法：normal_gmm；
- support_recursively_split：定义当前方法是否支持递归拆分；
- n_splits_candidates：指定生成聚合类数量的候选值；
- split_metric：指定在轮廓分数中定义无监督聚类度量的方式；
- covariance_type：指定使用的协方差参数类型；
- tol：控制提前终止迭代；
- reg_covar：添加非负正则化到协方差诊断中；
- n_init：要执行的初始化数。



5.7.6.4 使用说明

在数据建模流程中，normal_gmm 节点一般位于输出量为 DTable 的节点下游，模型（知识）聚合节点的上游，用于将数据根据内在性质分成不同的聚合类，进而针对特定聚合类作进一步分析。

5.8 数据降维

5.8.1 PCA

PCA 节点提供线性降维功能，可以将原高维空间中的数据点映射到低维度的空间中。其将原始线性相关变量表示的数据，利用正交变换成由若干个线性无关的新变量表示的数据。由于主成分的个数通常比原始维度少，所以主成分分析可以起到降低维度的效果。

5.8.1.1 输入量

PCA 节点输入量如表 45 所示。

表45 PCA 节点输入量列表

名称	类型	描述
in_data	DTable	需对变量特征进行降维的数据

5.8.1.2 输出量

PCA 节点输出量如表 46 所示。

表46 PCA 节点输出量列表

名称	类型	描述
out_data	DTable	在原始数据基础上新增降维后的特征变量
model	Model	降维后得到的模型



5.8.1.3 配置参数

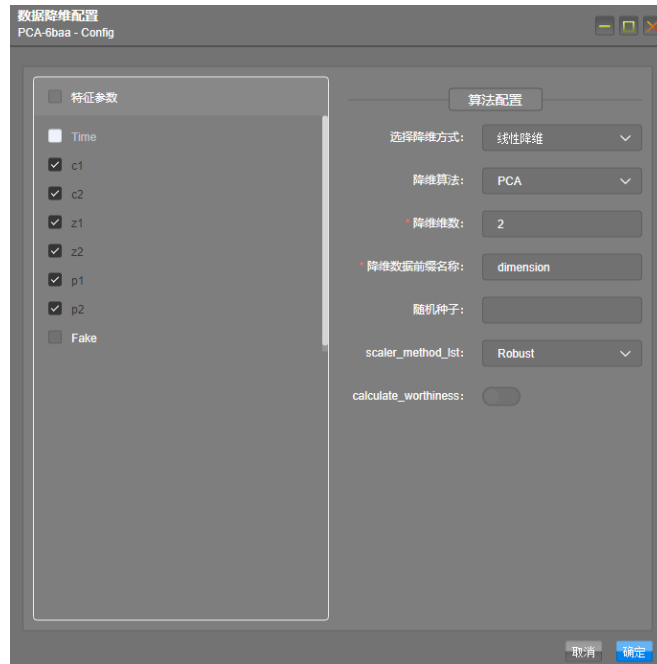


图203 PCA 节点属性

PCA 节点配置参数如图 203 所示，各参数描述如下：

- 节点名称：节点在建模流程图中的名称标注；
- 节点描述：节点描述信息，用于备注节点所实现的功能；
- 特征参数列表：勾选数据集内降维之前的特征参数，默认不会勾选输出参数；
- 降维方式：线性降维；
- 降维算法：PCA；
- 降维维数：降维后新生成的特征参数数量；
- 降维数据前缀名称：指定降维后新生成的降维数据前缀名称；
- 随机种子：可以为空或输入正整数，后者用于指定算法超参数初始值生成的规则；
- 数据变换：下拉选择数据变换方式。



5.8.1.4 使用说明

在数据建模流程中，PCA 节点一般位于输出量为 DTable 的节点下游，模型（知识）聚合节点的上游，用于筛除冗余信息或噪音、提高模型精度、寻找数据内部的本质结构特征等。

5.8.2 batch_PCA

batch_PCA 节点提供线性降维功能，可以将原高维空间中的数据点映射到低维度的空间中。基于一个最小 batch 的样本子集，每次取一个数据点作为样本增量，用每个新点更新其对期望特征向量的估计。这个方法比 PCA 计算速度更快。

5.8.2.1 输入量

batch_PCA 节点输入量如表 47 所示。

表47 batch_PCA 节点输入量列表

名称	类型	描述
in_data	DTable	需对变量特征进行降维的数据

5.8.2.2 输出量

batch_PCA 节点输出量如表 48 所示。

表48 batch_PCA 节点输出量列表

名称	类型	描述
out_data	DTable	在原始数据基础上新增降维后的特征变量
model	Model	降维后得到的模型



5.8.2.3 配置参数

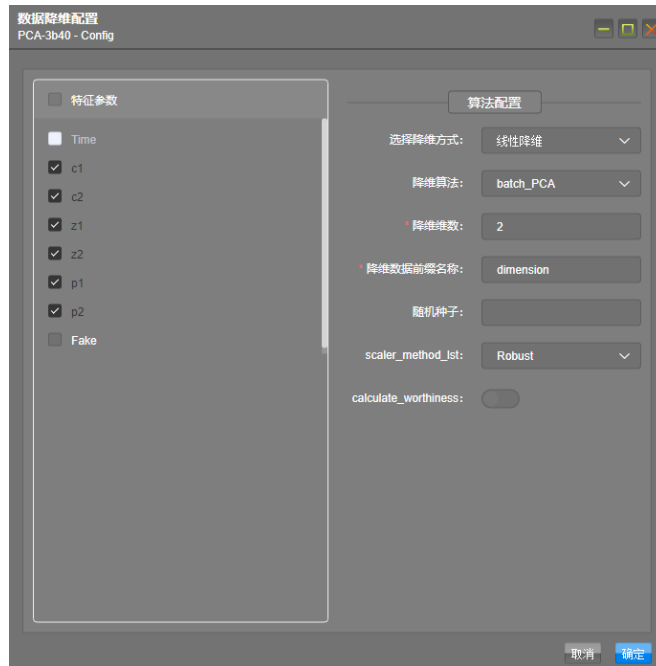


图204 batch_PCA 节点属性

batch_PCA 节点配置参数如图 204 所示，各参数描述如下：

- 节点名称：节点在建模流程图中的名称标注；
- 节点描述：节点描述信息，用于备注节点所实现的功能；
- 特征参数列表：勾选数据集内降维之前的特征参数；
- 降维方式：线性降维；
- 降维算法：batch_PCA；
- 降维维数：降维后新生成的特征参数数量；
- 降维数据前缀名称：指定降维后新生成的降维数据前缀名称；
- 随机种子：可以为空或输入正整数，后者用于指定算法超参数初始值生成的规则；
- 数据变换：下拉选择数据变换方式。

5.8.2.4 使用说明

在数据建模流程中，batch_PCA 节点一般位于输出量为 DTable 的节点下游，模型（知识）聚合节点的上游，用于筛除冗余信息或噪音、提高模型精度、寻找数据内部的本质结构特征等。



5.8.3 poly_KPCA

poly_KPCA 节点提供线性降维功能，可以将原高维空间中的数据点映射到低维度的空间中。其使用 poly 核函数的方法，将原始线性不可分的数据集投影到高维线性空间，再使用主成分分析的方法进行降维。

5.8.3.1 输入量

poly_KPCA 节点输入量如表 49 所示。

表49 poly_KPCA 节点输入量列表

名称	类型	描述
in_data	DTTable	需对变量特征进行降维的数据

5.8.3.2 输出量

poly_KPCA 节点输出量如表 50 所示。

表50 poly_KPCA 节点输出量列表

名称	类型	描述
out_data	DTTable	在原始数据基础上新增降维后的特征变量
model	Model	降维后得到的模型

5.8.3.3 配置参数

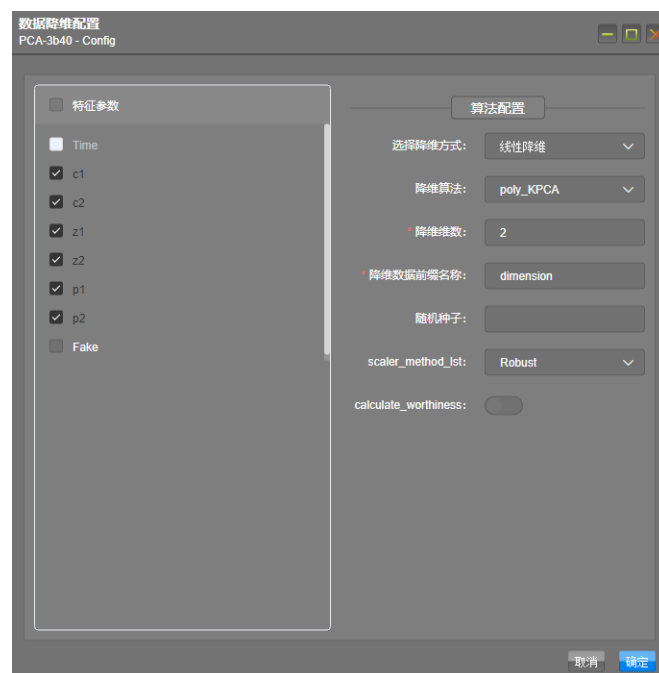


图205 poly_KPCA 节点属性

poly_KPCA 节点配置参数如图 205 所示，各参数描述如下：

- 节点名称：节点在建模流程图中的名称标注；
- 节点描述：节点描述信息，用于备注节点所实现的功能；
- 特征参数列表：勾选数据集内降维之前的特征参数；
- 降维方式：线性降维；
- 降维算法：poly_KPCA；
- 降维维数：降维后新生成的特征参数数量；
- 降维数据前缀名称：指定降维后新生成的降维数据前缀名称；
- 随机种子：可以为空或输入正整数，后者用于指定算法超参数初始值生成的规则；
- 数据变换：下拉选择数据变换方式。

5.8.3.4 使用说明

在数据建模流程中，poly_KPCA 节点一般位于输出量为 DTable 的节点下游，模型（知识）聚合节点的上游，用于筛除冗余信息或噪音、提高模型精度、寻找数据内部的本质结构特征等。

5.8.4 rbf_KPCA

rbf_KPCA 节点提供线性降维功能，可以将原高维空间中的数据点映射到低维度的空间中。其使用 rbf 核函数的方法，将原始线性不可分的数据集投影到高维线性空间，再使用主成分分析的方法进行降维。

5.8.4.1 输入量

rbf_KPCA 节点输入量如表 51 所示。

表51 rbf_KPCA 节点输入量列表

名称	类型	描述
in_data	DTable	需对变量特征进行降维的数据

5.8.4.2 输出量

rbf_KPCA 节点输出量如表 52 所示。

表52 PCA 节点输出量列表



名称	类型	描述
out_data	DTTable	在原始数据基础上新增降维后的特征变量
model	Model	降维后得到的模型

5.8.4.3 配置参数

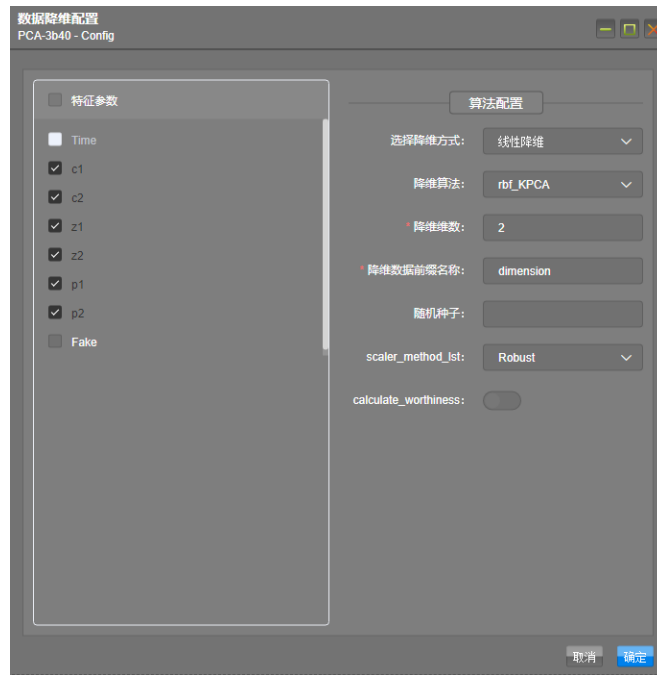


图206 rbf_KPCA 节点属性

rbf_KPCA 节点配置参数如图 206 所示，各参数描述如下：

- 节点名称：节点在建模流程图中的名称标注；
- 节点描述：节点描述信息，用于备注节点所实现的功能；
- 特征参数列表：勾选数据集内降维之前的特征参数；
- 降维方式：线性降维；
- 降维算法：rbf_KPCA；
- 降维维数：降维后新生成的特征参数数量；
- 降维数据前缀名称：指定降维后新生成的降维数据前缀名称；
- 随机种子：可以为空或输入正整数，后者用于指定算法超参数初始值生成的规则；
- 数据变换：下拉选择数据变换方式。



5.8.4.4 使用说明

在数据建模流程中,rbf_KPCA 节点一般位于输出量为 DTable 的节点下游,模型(知识)聚合节点的上游,用于筛除冗余信息或噪音、提高模型精度、寻找数据内部的本质结构特征等。

5.8.5 sigmoid_KPCA

sigmoid_KPCA 节点提供线性降维功能,可以将原高维空间中的数据点映射到低维度的空间中。其使用 sigmoid 核函数的方法,将原始线性不可分的数据集投影到高维线性空间,再使用主成分分析的方法进行降维。

5.8.5.1 输入量

sigmoid_KPCA 节点输入量如表 53 所示。

表53 sigmoid_KPCA 节点输入量列表

名称	类型	描述
in_data	DTable	需对变量特征进行降维的数据

5.8.5.2 输出量

sigmoid_KPCA 节点输出量如表 54 所示。

表54 sigmoid_KPCA 节点输出量列表

名称	类型	描述
out_data	DTable	在原始数据基础上新增降维后的特征变量
model	Model	降维后得到的模型



5.8.5.3 配置参数

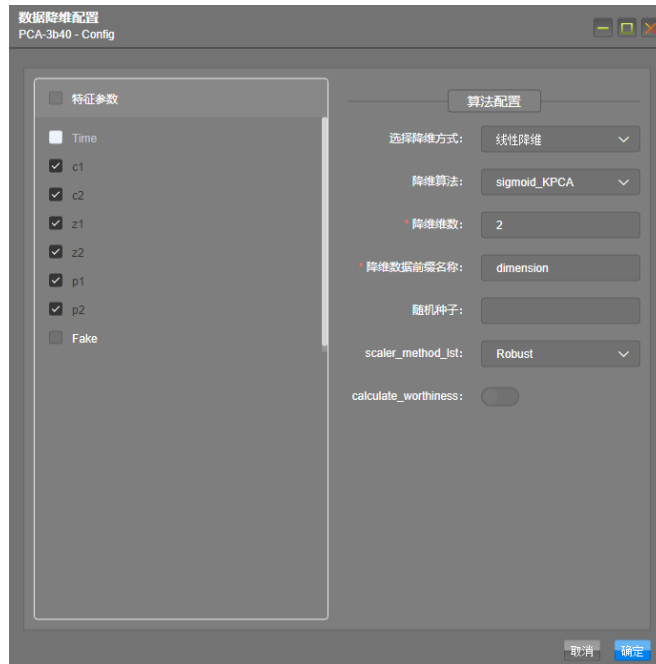


图207 sigmoid_KPCA 节点属性

sigmoid_KPCA 节点配置参数如图 207 所示，各参数描述如下：

- 节点名称：节点在建模流程图中的名称标注；
- 节点描述：节点描述信息，用于备注节点所实现的功能；
- 特征参数列表：勾选数据集内降维之前的特征参数；
- 降维方式：线性降维；
- 降维算法：sigmoid_KPCA；
- 降维维数：降维后新生成的特征参数数量；
- 降维数据前缀名称：指定降维后新生成的降维数据前缀名称；
- 随机种子：可以为空或输入正整数，后者用于指定算法超参数初始值生成的规则；
- 数据变换：下拉选择数据变换方式。

5.8.5.4 使用说明

在数据建模流程中，sigmoid_KPCA 节点一般位于输出量为 DTable 的节点下游，模型（知识）聚合节点的上游，用于筛除冗余信息或噪音、提高模型精度、寻找数据内部的本质结构特征等。



5.8.6 cosine_KPCA

cosine_KPCA 节点提供线性降维功能, 可以将原高维空间中的数据点映射到低维度的空间中。其使用 cosine 核函数的方法, 将原始线性不可分的数据集投影到高维线性空间, 再使用主成分分析的方法进行降维。

5.8.6.1 输入量

cosine_KPCA 节点输入量如表 55 所示。

表55 cosine_KPCA 节点输入量列表

名称	类型	描述
in_data	DTTable	需对变量特征进行降维的数据

5.8.6.2 输出量

cosine_KPCA 节点输出量如表 56 所示。

表56 cosine_KPCA 节点输出量列表

名称	类型	描述
out_data	DTTable	在原始数据基础上新增降维后的特征变量
model	Model	降维后得到的模型

5.8.6.3 配置参数

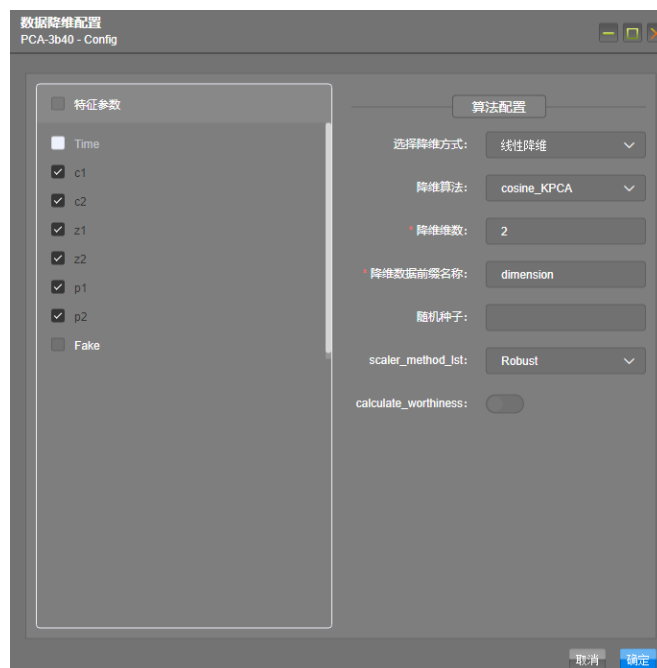


图208 cosine_KPCA 节点属性



cosine_KPCA 节点配置参数如图 208 所示，各参数描述如下：

- 节点名称：节点在建模流程图中的名称标注；
- 节点描述：节点描述信息，用于备注节点所实现的功能；
- 特征参数列表：勾选数据集内降维之前的特征参数；
- 降维方式：线性降维；
- 降维算法：cosine_KPCA；
- 降维维数：降维后新生成的特征参数数量；
- 降维数据前缀名称：指定降维后新生成的降维数据前缀名称；
- 随机种子：可以为空或输入正整数，后者用于指定算法超参数初始值生成的规则；
- 数据变换：下拉选择数据变换方式。

5.8.6.4 使用说明

在数据建模流程中，cosine_KPCA 节点一般位于输出量为 DTable 的节点下游，模型（知识）聚合节点的上游，用于筛除冗余信息或噪音、提高模型精度、寻找数据内部的本质结构特征等。

5.8.7 FA

FA 节点提供线性降维功能，可以将原高维空间中的数据点映射到低维度的空间中。FA（factor analysis，因子分析）假设原数据集可以写成多个因子的线性组合，即 $Y=AX$ 。因子分析通过主成分分析、最大似然或因式分解等计算方式计算载荷矩阵 A。再通过旋转因子的方法计算得分矩阵与降维因子 X。

5.8.7.1 输入量

FA 节点输入量如表 57 所示。

表57 FA 节点输入量列表

名称	类型	描述
in_data	DTable	需对变量特征进行降维的数据

5.8.7.2 输出量

FA 节点输出量如表 58 所示。



表58 FA 节点输出量列表

名称	类型	描述
out_data	DTTable	在原始数据基础上新增降维后的特征变量
model	Model	降维后得到的模型

5.8.7.3 配置参数

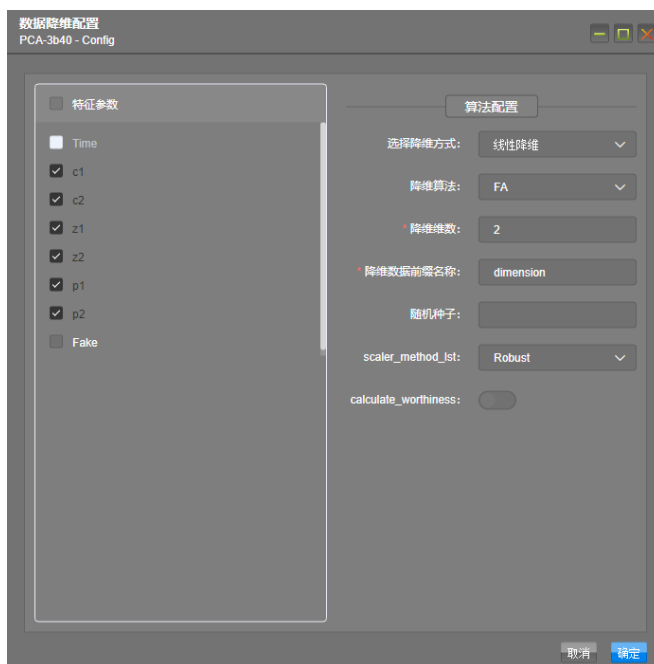


图209 FA 节点属性

FA 节点配置参数如图 209 所示，各参数描述如下：

- 节点名称：节点在建模流程图中的名称标注；
- 节点描述：节点描述信息，用于备注节点所实现的功能；
- 特征参数列表：勾选数据集内降维之前的特征参数；
- 降维方式：线性降维；
- 降维算法：FA；
- 降维维数：降维后新生成的特征参数数量；
- 降维数据前缀名称：指定降维后新生成的降维数据前缀名称；
- 随机种子：可以为空或输入正整数，后者用于指定算法超参数初始值生成的规则；
- 数据变换：下拉选择数据变换方式。



5.8.7.4 使用说明

在数据建模流程中，FA 节点一般位于输出量为 DTable 的节点下游，模型（知识）聚合节点的上游，用于筛除冗余信息或噪音、提高模型精度、寻找数据内部的本质结构特征等。

5.8.8 DL

DL 节点提供线性降维功能，可以将原高维空间中的数据点映射到低维度的空间中。DL（Dictionary Learning）是一种使用稀疏模型、最小化损失函数方法对原始数据集进行降维的方法。稀疏模型将大量的冗余变量去除，只保留与响应变量最相关的解释变量，简化了模型的同时却保留了数据集中最重要的信息，有效地解决了高维数据集建模中的诸多问题。

5.8.8.1 输入量

DL 节点输入量如表 59 所示。

表59 DL 节点输入量列表

名称	类型	描述
in_data	DTable	需对变量特征进行降维的数据

5.8.8.2 输出量

DL 节点输出量如表 60 所示。

表60 DL 节点输出量列表

名称	类型	描述
out_data	DTable	在原始数据基础上新增降维后的特征变量
model	Model	降维后得到的模型



5.8.8.3 配置参数

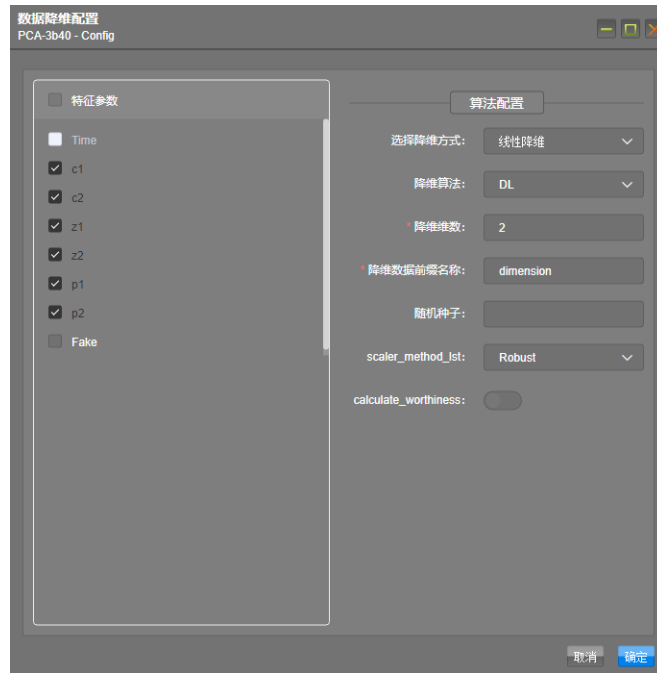


图210 DL 节点属性

DL 节点配置参数如图 210 所示，各参数描述如下：

- 节点名称：节点在建模流程图中的名称标注；
- 节点描述：节点描述信息，用于备注节点所实现的功能；
- 特征参数列表：勾选数据集内降维之前的特征参数；
- 降维方式：线性降维；
- 降维算法：DL；
- 降维维数：降维后新生成的特征参数数量；
- 降维数据前缀名称：指定降维后新生成的降维数据前缀名称；
- 随机种子：可以为空或输入正整数，后者用于指定算法超参数初始值生成的规则；
- 数据变换：下拉选择数据变换方式。

5.8.8.4 使用说明

在数据建模流程中，DL 节点一般位于输出量为 DTable 的节点下游，模型（知识）聚合节点的上游，用于筛除冗余信息或噪音、提高模型精度、寻找数据内部的本质结构特征等。



5.8.9 LSA

LSA 节点提供线性降维功能，可以将原高维空间中的数据点映射到低维度的空间中。LSA 节点基于截断奇异值分解（TruncatedSVD），提前设定分解成分的个数，可以更有效的处理稀疏矩阵。

5.8.9.1 输入量

LSA 节点输入量如表 61 所示。

表61 LSA 节点输入量列表

名称	类型	描述
in_data	DTTable	需对变量特征进行降维的数据

5.8.9.2 输出量

LSA 节点输出量如表 62 所示。

表62 LSA 节点输出量列表

名称	类型	描述
out_data	DTTable	在原始数据基础上新增降维后的特征变量
model	Model	降维后得到的模型

5.8.9.3 配置参数

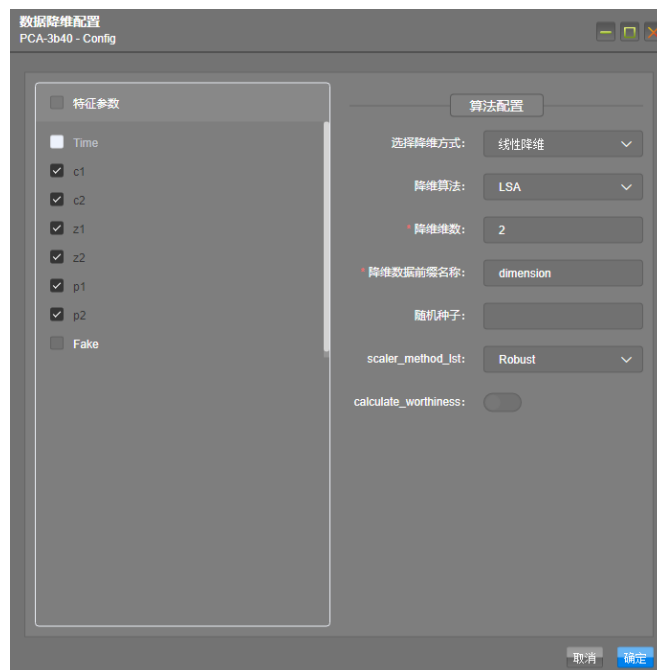


图211 LSA 节点属性



LSA 节点配置参数如图 211 所示，各参数描述如下：

- 节点名称：节点在建模流程图中的名称标注；
- 节点描述：节点描述信息，用于备注节点所实现的功能；
- 特征参数列表：勾选数据集内降维之前的特征参数；
- 降维方式：线性降维；
- 降维算法：LSA；
- 降维维数：降维后新生成的特征参数数量；
- 降维数据前缀名称：指定降维后新生成的降维数据前缀名称；
- 随机种子：可以为空或输入正整数，后者用于指定算法超参数初始值生成的规则；
- 数据变换：下拉选择数据变换方式。

5.8.9.4 使用说明

在数据建模流程中，LSA 节点一般位于输出量为 DTTable 的节点下游，模型（知识）聚合节点的上游，用于筛除冗余信息或噪音、提高模型精度、寻找数据内部的本质结构特征等。

5.8.10 ICA

ICA 节点提供线性降维功能，可以将原高维空间中的数据点映射到低维度的空间中。ICA（Independent Component Analysis，独立成分分析）是一种将观测信号分解为加性的原始信号的一种方法，它通过假设原始信号是相互独立且非高斯分布的，从而在观测信号中分离原始信号。

5.8.10.1 输入量

ICA 节点输入量如表 63 所示。

表63 ICA 节点输入量列表

名称	类型	描述
in_data	DTTable	需对变量特征进行降维的数据

5.8.10.2 输出量

ICA 节点输出量如表 64 所示。



表64 ICA 节点输出量列表

名称	类型	描述
out_data	DTTable	在原始数据基础上新增降维后的特征变量
model	Model	降维后得到的模型

5.8.10.3 配置参数

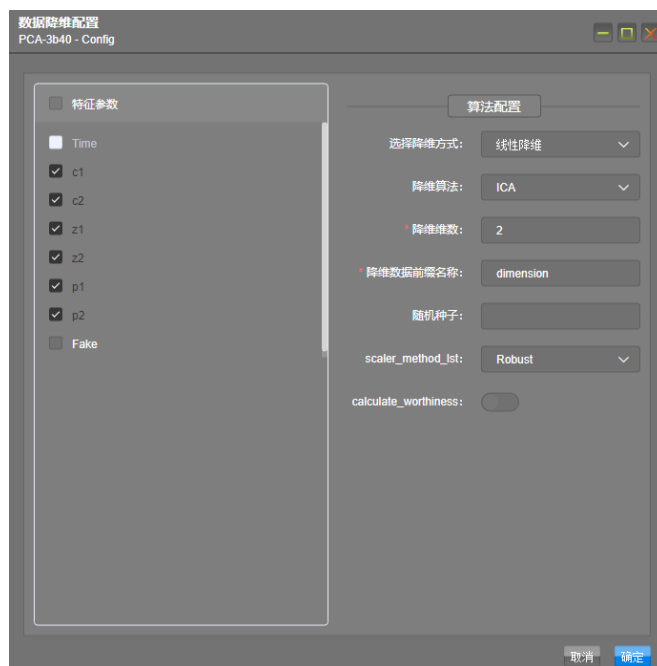


图212 ICA 节点属性

ICA 节点配置参数如图 212 所示，各参数描述如下：

- 节点名称：节点在建模流程图中的名称标注；
- 节点描述：节点描述信息，用于备注节点所实现的功能；
- 特征参数列表：勾选数据集内降维之前的特征参数；
- 降维方式：线性降维；
- 降维算法：ICA；
- 降维维数：降维后新生成的特征参数数量；
- 降维数据前缀名称：指定降维后新生成的降维数据前缀名称；
- 随机种子：可以为空或输入正整数，后者用于指定算法超参数初始值生成的规则；
- 数据变换：下拉选择数据变换方式。



5.8.10.4 使用说明

在数据建模流程中，ICA 节点一般位于输出量为 DTable 的节点下游，模型（知识）聚合节点的上游，用于筛除冗余信息或噪音、提高模型精度、寻找数据内部的本质结构特征等。

5.8.11 isomap

isomap 节点提供非线性降维功能，可以将原高维空间中的数据点映射到低维度的空间中。这种方法是一种全局方法，需要计算任意两点间的测地距离，计算量大，但是它比 LLE 方法保留了更多的信息。

5.8.11.1 输入量

isomap 节点输入量如表 65 所示。

表65 isomap 节点输入量列表

名称	类型	描述
in_data	DTable	需对变量特征进行降维的数据

5.8.11.2 输出量

isomap 节点输出量如表 66 所示。

表66 isomap 节点输出量列表

名称	类型	描述
out_data	DTable	在原始数据基础上新增降维后的特征变量
model	Model	降维后得到的模型



5.8.11.3 配置参数

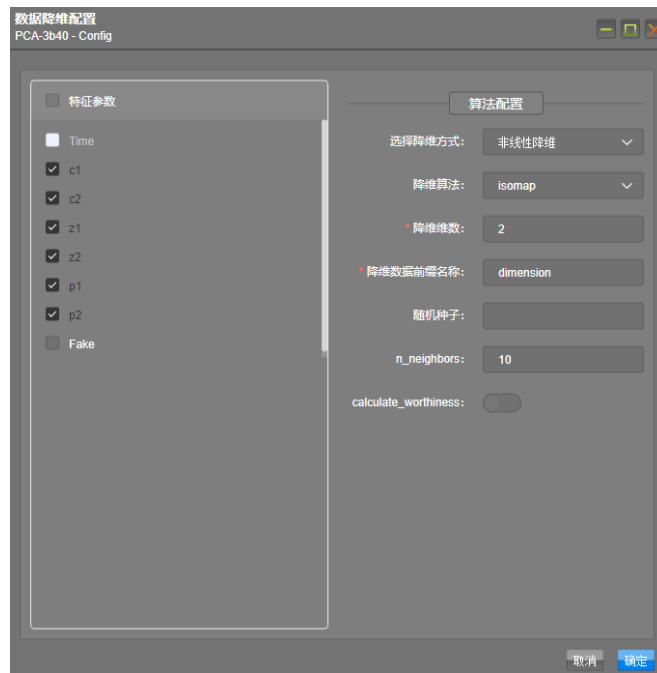


图213 isomap 节点属性

isomap 节点配置参数如图 213 所示，各参数描述如下：

- 节点名称：节点在建模流程图中的名称标注；
- 节点描述：节点描述信息，用于备注节点所实现的功能；
- 特征参数列表：勾选数据集内降维之前的特征参数；
- 降维方式：非线性降维；
- 降维算法：isomap；
- 降维维数：降维后新生成的特征参数数量；
- 降维数据前缀名称：指定降维后新生成的降维数据前缀名称；
- 随机种子：可以为空或输入正整数，后者用于指定算法超参数初始值生成的规则；
- n_neighbors：邻近点个数，需输入不超过 100 的正整数。

5.8.11.4 使用说明

在数据建模流程中，isomap 节点一般位于输出量为 DTable 的节点下游，模型（知识）聚合节点的上游，用于筛除冗余信息或噪音、提高模型精度、寻找数据内部的本质结构特征等。



5.8.12 LLE

LLE 节点提供非线性降维功能, 可以将原高维空间中的数据点映射到低维度的空间中。LLE (Locally Linear Embedding, 局部线性嵌入) 方法认为流形在局部上是线性的, 每个样本点可以由它的近邻点线性表示。所以, 这种方法通过保持近邻点权值不变, 从而在低维空间中找到对应的低维样本点。这种方法计算量小, 可以学习任意维的局部线性低维流形, 并且具有整体最优解, 不需迭代; 但是这种方法对噪音很敏感, 要求数据集必须稠密采样。

5.8.12.1 输入量

LLE 节点输入量如表 67 所示。

表67 LLE 节点输入量列表

名称	类型	描述
in_data	DTTable	需对变量特征进行降维的数据

5.8.12.2 输出量

LLE 节点输出量如表 68 所示。

表68 LLE 节点输出量列表

名称	类型	描述
out_data	DTTable	在原始数据基础上新增降维后的特征变量
model	Model	降维后得到的模型



5.8.12.3 配置参数

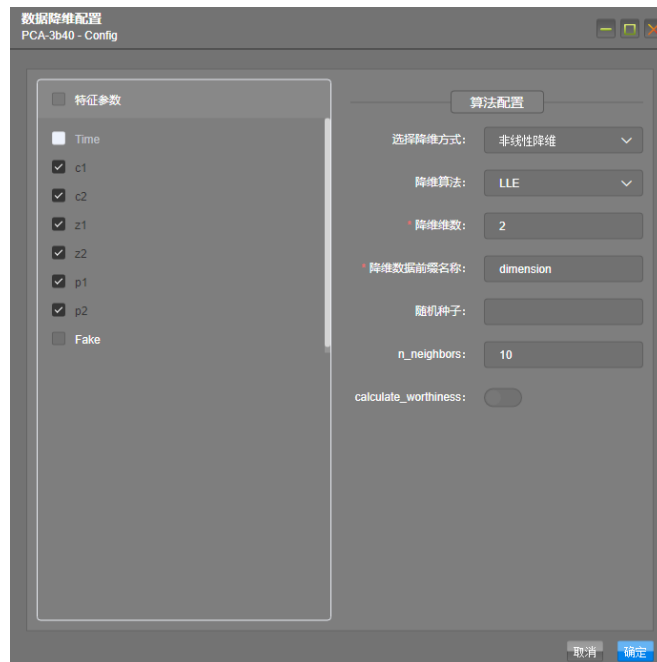


图214 LLE 节点属性

LLE 节点配置参数如图 214 所示，各参数描述如下：

- 节点名称：节点在建模流程图中的名称标注；
- 节点描述：节点描述信息，用于备注节点所实现的功能；
- 特征参数列表：勾选数据集内降维之前的特征参数；
- 降维方式：非线性降维；
- 降维算法：LLE；
- 降维维数：降维后新生成的特征参数数量；
- 降维数据前缀名称：指定降维后新生成的降维数据前缀名称；
- 随机种子：可以为空或输入正整数，后者用于指定算法超参数初始值生成的规则；
- n_neighbors：邻近点个数，需输入不超过 100 的正整数。

5.8.12.4 使用说明

在数据建模流程中，LLE 节点一般位于输出量为 DTable 的节点下游，模型（知识）聚合节点的上游，用于筛除冗余信息或噪音、提高模型精度、寻找数据内部的本质结构特征等。



5.8.13 LTSA

LTSA 节点提供非线性降维功能，可以将原高维空间中的数据点映射到低维度的空间中。在 LLE 算法的基础上，将局部线性嵌入方法替换为 LTSA（Linear Local Tangent Space Alignment）。

5.8.13.1 输入量

LTSA 节点输入量如表 69 所示。

表69 LTSA 节点输入量列表

名称	类型	描述
in_data	DTTable	需对变量特征进行降维的数据

5.8.13.2 输出量

LTSA 节点输出量如表 70 所示。

表70 LTSA 节点输出量列表

名称	类型	描述
out_data	DTTable	在原始数据基础上新增降维后的特征变量
model	Model	降维后得到的模型

5.8.13.3 配置参数

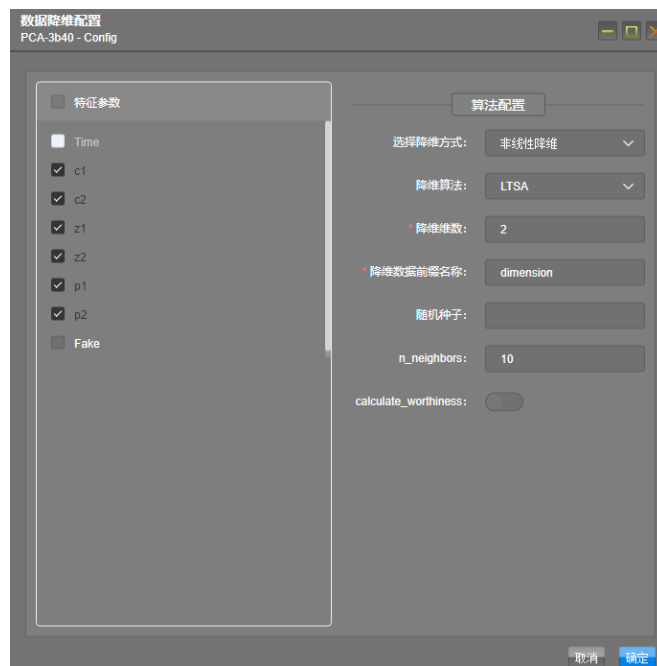


图215 LTSA 节点属性



LTSA 节点配置参数如图 215 所示，各参数描述如下：

- 节点名称：节点在建模流程图中的名称标注；
- 节点描述：节点描述信息，用于备注节点所实现的功能；
- 特征参数列表：勾选数据集内降维之前的特征参数；
- 降维方式：非线性降维；
- 降维算法：LTSA；
- 降维维数：降维后新生成的特征参数数量；
- 降维数据前缀名称：指定降维后新生成的降维数据前缀名称；
- 随机种子：可以为空或输入正整数，后者用于指定算法超参数初始值生成的规则；
- n_neighbors：邻近点个数，需输入不超过 100 的正整数。

5.8.13.4 使用说明

在数据建模流程中，LTSA 节点一般位于输出量为 DTable 的节点下游，模型（知识）聚合节点的上游，用于筛除冗余信息或噪音、提高模型精度、寻找数据内部的本质结构特征等。

5.8.14 hessian_LLE

hessian_LLE 节点提供非线性降维功能，可以将原高维空间中的数据点映射到低维度的空间中。在 LLE 算法的基础上，将局部线性嵌入方法替换为 hessian_LLE（Hessian Locally Linear Embedding）。

5.8.14.1 输入量

hessian_LLE 节点输入量如表 71 所示。

表71 hessian_LLE 节点输入量列表

名称	类型	描述
in_data	DTable	需对变量特征进行降维的数据

5.8.14.2 输出量

hessian_LLE 节点输出量如表 72 所示。

表72 hessian_LLE 节点输出量列表



名称	类型	描述
out_data	DTTable	在原始数据基础上新增降维后的特征变量
model	Model	降维后得到的模型

5.8.14.3 配置参数

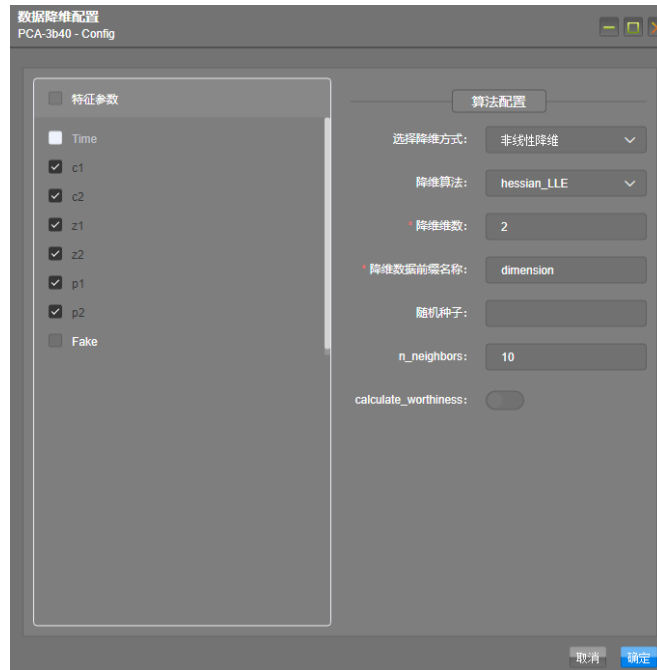


图216 hessian_LLE 节点属性

hessian_LLE 节点配置参数如图 216 所示，各参数描述如下：

- 节点名称：节点在建模流程图中的名称标注；
- 节点描述：节点描述信息，用于备注节点所实现的功能；
- 特征参数列表：勾选数据集内降维之前的特征参数；
- 降维方式：非线性降维；
- 降维算法：hessian_LLE；
- 降维维数：降维后新生成的特征参数数量；
- 降维数据前缀名称：指定降维后新生成的降维数据前缀名称；
- 随机种子：可以为空或输入正整数，后者用于指定算法超参数初始值生成的规则；
- n_neighbors：邻近点个数，需输入不超过 100 的正整数。



5.8.14.4 使用说明

在数据建模流程中，hessian_LLE 节点一般位于输出量为 DTable 的节点下游，模型（知识）聚合节点的上游，用于筛除冗余信息或噪音、提高模型精度、寻找数据内部的本质结构特征等。

5.8.15 modified_LLE

modified_LLE 节点提供非线性降维功能，可以将原高维空间中的数据点映射到低维度的空间中。在 LLE 算法的基础上，将局部线性嵌入方法替换为 modified_LLE（Modified Locally Linear Embedding）。

5.8.15.1 输入量

modified_LLE 节点输入量如表 73 所示。

表73 modified_LLE 节点输入量列表

名称	类型	描述
in_data	DTable	需对变量特征进行降维的数据

5.8.15.2 输出量

modified_LLE 节点输出量如表 74 所示。

表74 modified_LLE 节点输出量列表

名称	类型	描述
out_data	DTable	在原始数据基础上新增降维后的特征变量
model	Model	降维后得到的模型



5.8.15.3 配置参数

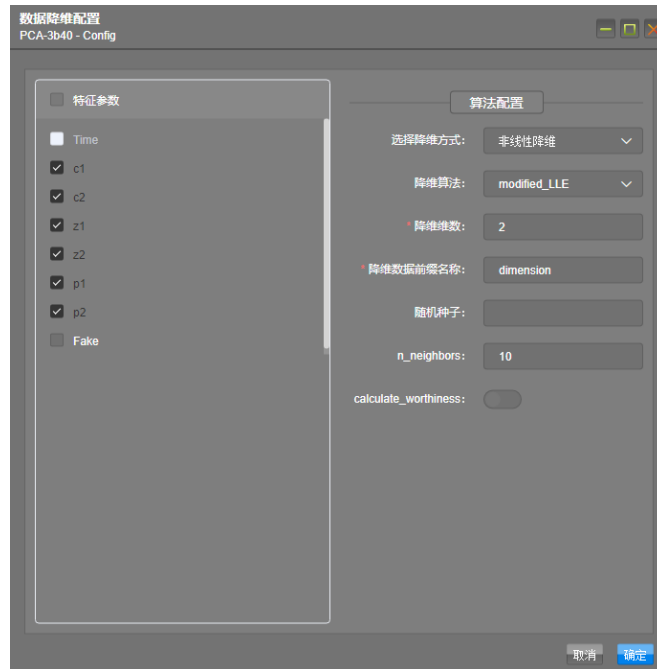


图217 modified_LLE 节点属性

modified_LLE 节点配置参数如图 217 所示，各参数描述如下：

- 节点名称：节点在建模流程图中的名称标注；
- 节点描述：节点描述信息，用于备注节点所实现的功能；
- 特征参数列表：勾选数据集内降维之前的特征参数；
- 降维方式：非线性降维；
- 降维算法：modified_LLE；
- 降维维数：降维后新生成的特征参数数量；
- 降维数据前缀名称：指定降维后新生成的降维数据前缀名称；
- 随机种子：可以为空或输入正整数，后者用于指定算法超参数初始值生成的规则；
- n_neighbors：邻近点个数，需输入不超过 100 的正整数。

5.8.15.4 使用说明

在数据建模流程中，modified_LLE 节点一般位于输出量为 DTable 的节点下游，模型（知识）聚合节点的上游，用于筛除冗余信息或噪音、提高模型精度、寻找数据内部的本质结构特征等。



5.9 数据变换

5.9.1 MinMax

MinMax 将原始变量值线性缩放至指定区间。

5.9.1.1 输入量

MinMax 节点输入量如表 75 所示。

表75 MinMax 节点输入量列表

名称	类型	描述
in_data	DTTable	需要进行数据变换的数据

5.9.1.2 输出量

MinMax 节点输出量如表 76 所示。

表76 MinMax 节点输出量列表

名称	类型	描述
out_data	DTTable	完成数据变换的数据
model	Model	生成新特征的映射关系模型

5.9.1.3 配置参数



图218 MinMax 节点属性

MinMax 节点配置参数如图 218 所示，各参数描述如下：

- 节点名称：节点在建模流程图中的名称标注；



- 节点描述：节点描述信息，用于备注节点所实现的功能；
- 选择算法：MinMax；
- feature_range：缩放后变量区间。

需要注意的是，当该节点未执行时，原始特征量对于该节点是未知的。当该节点被执行时，系统会自动弹出数据变换窗口，用户在弹窗中对新特征生成的规则进行编辑，如图 219 所示。

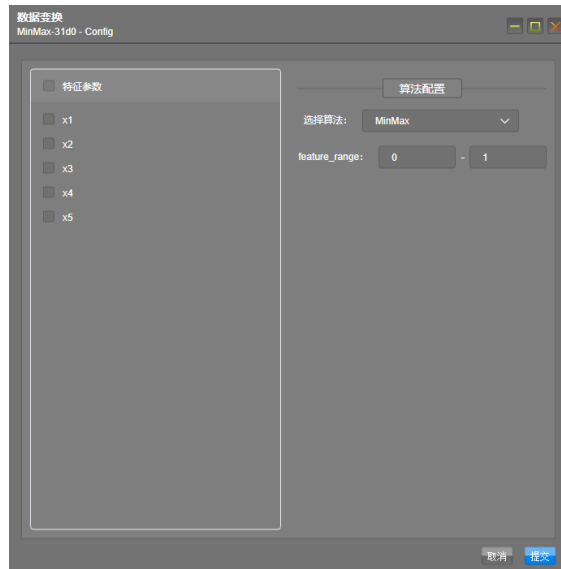


图219 MinMax 配置窗口

如图 219 所示，MinMax 配置弹窗左侧为原始特征名称列表，右侧为用户指定数据变换的规则。

5.9.1.4 使用说明

在数据建模流程中，MinMax 节点一般位于输出量为 DTable 节点下游，与接收 DTable 作为训练数据的模型训练节点之间。

5.9.2 MaxAbs

MaxAbs 将原始变量值中绝对值最大的数据缩放为 1（或-1），其余数据按相同比例进行缩放。

5.9.2.1 输入量

MaxAbs 节点输入量如表 77 所示。

表77 MaxAbs 节点输入量列表



名称	类型	描述
in_data	DTable	需要进行数据变换的数据

5.9.2.2 输出量

MaxAbs 节点输出量如表 78 所示。

表78 MaxAbs 节点输出量列表

名称	类型	描述
out_data	DTable	完成数据变换的数据
model	Model	生成新特征的映射关系模型

5.9.2.3 配置参数

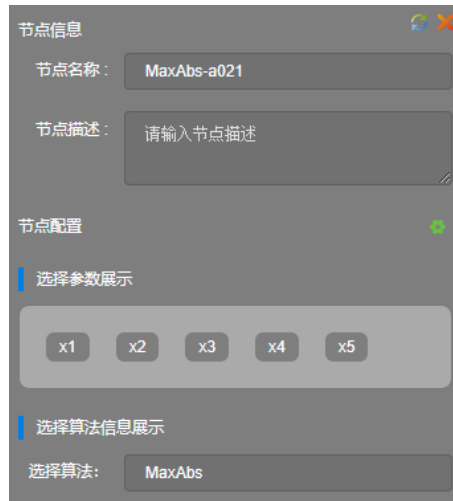


图220 MaxAbs 节点属性

MaxAbs 节点配置参数如图 220 所示，各参数描述如下：

- 节点名称：节点在建模流程图中的名称标注；
- 节点描述：节点描述信息，用于备注节点所实现的功能；
- 选择算法：MaxAbs。

需要注意的是，当该节点未执行时，原始特征量对于该节点是未知的。当该节点被执行时，系统会自动弹出数据变换窗口，用户在弹窗中对新特征生成的规则进行编辑，如图 221 所示。





图221 MaxAbs 配置窗口

如图 221 所示，MaxAbs 配置弹窗左侧为原始特征名称列表，右侧为用户指定数据变换的规则。

5.9.2.4 使用说明

在数据建模流程中，MaxAbs 节点一般位于输出量为 DTable 节点下游，与接收 DTable 作为训练数据的模型训练节点之间。

5.9.3 Z-score

Z-score 使用均值和方差对数据标准化，把数据投影到均值为 0，方差为 1 的分布上。

5.9.3.1 输入量

Z-score 节点输入量如表 79 所示。

表79 Z-score 节点输入量列表

名称	类型	描述
in_data	DTable	需要进行数据变换的数据

5.9.3.2 输出量

Z-score 节点输出量如表 80 所示。

表80 Z-score 节点输出量列表

名称	类型	描述
out_data	DTable	完成数据变换的数据



5.9.3.3 配置参数



图222 Z-score 节点属性

Z-score 节点配置参数如图 222 所示，各参数描述如下：

- 节点名称：节点在建模流程图中的名称标注；
- 节点描述：节点描述信息，用于备注节点所实现的功能；
- 选择算法：Z-score。

需要注意的是，当该节点未执行时，原始特征量对于该节点是未知的。当该节点被执行时，系统会自动弹出数据变换窗口，用户在弹窗中对新特征生成的规则进行编辑，如图 223 所示。

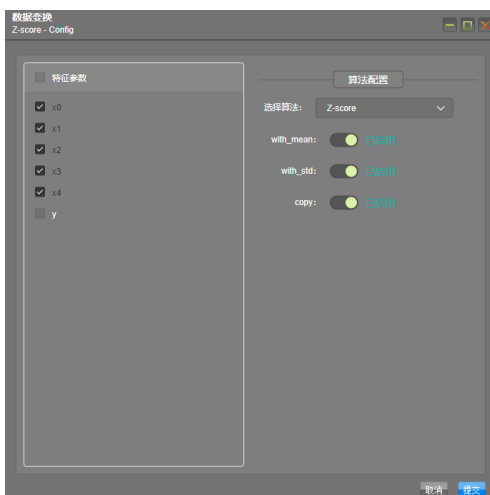


图223 Z-score 配置窗口

如图 223 所示，Z-score 配置弹窗左侧为原始特征名称列表，右侧为用户指定数据变换的规则。



5.9.3.4 使用说明

在数据建模流程中，Z-score 节点一般位于输出量为 DTable 节点下游，与接收 DTable 作为训练数据的模型训练节点之间。

5.9.4 Robust

Robust 使用原始变量的中位数和四分位数进行缩放，以确保每个变换后特征的统计属性都位于同一范围。

5.9.4.1 输入量

Robust 节点输入量如表 81 所示。

表81 Robust 节点输入量列表

名称	类型	描述
in_data	DTable	需要进行数据变换的数据

5.9.4.2 输出量

Robust 节点输出量如表 82 所示。

表82 Robust 节点输出量列表

名称	类型	描述
out_data	DTable	完成数据变换的数据
model	Model	生成新特征的映射关系模型



5.9.4.3 配置参数

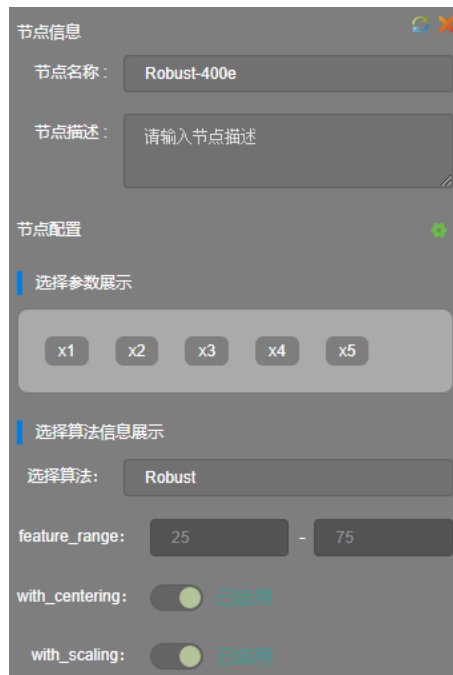


图224 Robust 节点属性

Robust 节点配置参数如图 224 所示，各参数描述如下：

- 节点名称：节点在建模流程图中的名称标注；
- 节点描述：节点描述信息，用于备注节点所实现的功能；
- 选择算法：Robust；
- feature_range：统计属性范围；
- with_centering：启用状态下，将数据在缩放前居中；
- with_scaling：启用状态下，将数据缩放到四分位范围。

需要注意的是，当该节点未执行时，原始特征量对于该节点是未知的。当该节点被执行时，系统会自动弹出数据变换窗口，用户在弹窗中对新特征生成的规则进行编辑，如图 225 所示。



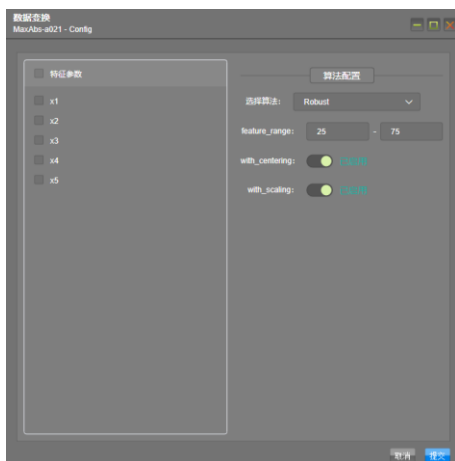


图225 Robust 配置窗口

如图 225 所示，Robust 配置弹窗左侧为原始特征名称列表，右侧为用户指定数据变换的规则。

5.9.4.4 使用说明

在数据建模流程中，Robust 节点一般位于输出量为 DTable 节点下游，与接收 DTable 作为训练数据的模型训练节点之间。

5.9.5 Multiplication

Multiplication 将原始变量值按倍数放大。

5.9.5.1 输入量

Multiplication 节点输入量如表 83 所示。

表83 Multiplication 节点输入量列表

名称	类型	描述
in_data	DTable	需要进行数据变换的数据

5.9.5.2 输出量

Multiplication 节点输出量如表 84 所示。

表84 Multiplication 节点输出量列表

名称	类型	描述
out_data	DTable	完成数据变换的数据
model	Model	生成新特征的映射关系模型



5.9.5.3 配置参数



图226 Multiplication 节点属性

Multiplication 节点配置参数如图 226 所示，各参数描述如下：

- 节点名称：节点在建模流程图中的名称标注；
- 节点描述：节点描述信息，用于备注节点所实现的功能；
- 选择算法：Multiplication；
- 放大倍数：变量放大的倍数。

需要注意的是，当该节点未执行时，原始特征量对于该节点是未知的。当该节点被执行时，系统会自动弹出数据变换窗口，用户在弹窗中对新特征生成的规则进行编辑，如图 227 所示。



图227 Multiplication 配置窗口



如图 227 所示，Multiplication 配置弹窗左侧为原始特征名称列表，右侧为用户指定数据变换的规则。

5.9.5.4 使用说明

在数据建模流程中，Multiplication 节点一般位于输出量为 DTable 节点下游，与接收 DTable 作为训练数据的模型训练节点之间。

5.9.6 Norm-1

Norm-1 将每个样本独立地根据其 L1 范数来归一化。

5.9.6.1 输入量

Norm-1 节点输入量如表 85 所示。

表85 Norm-1 节点输入量列表

名称	类型	描述
in_data	DTable	需要进行数据变换的数据

5.9.6.2 输出量

Norm-1 节点输出量如表 86 所示。

表86 Norm-1 节点输出量列表

名称	类型	描述
out_data	DTable	完成数据变换的数据
model	Model	生成新特征的映射关系模型

5.9.6.3 配置参数

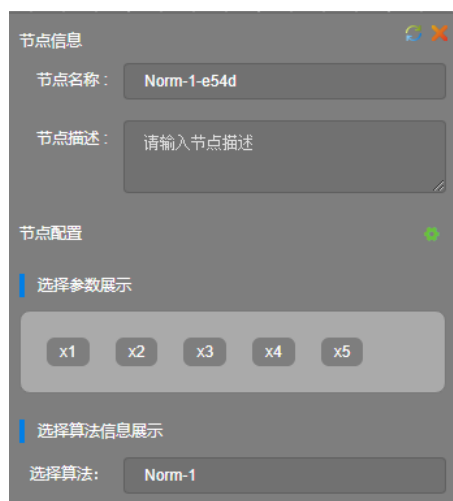


图228 Norm-1 节点属性

Norm-1 节点配置参数如图 228 所示，各参数描述如下：

- 节点名称：节点在建模流程图中的名称标注；
- 节点描述：节点描述信息，用于备注节点所实现的功能；
- 选择算法：Norm-1；。

需要注意的是，当该节点未执行时，原始特征量对于该节点是未知的。当该节点被执行时，系统会自动弹出数据变换窗口，用户在弹窗中对新特征生成的规则进行编辑，如图 229 所示。



图229 Norm-1 配置窗口

如图 229 所示，Norm-1 配置弹窗左侧为原始特征名称列表，右侧为用户指定数据变换的规则。

5.9.6.4 使用说明

在数据建模流程中，Norm-1 节点一般位于输出量为 DTable 节点下游，与接收 DTable 作为训练数据的模型训练节点之间。

5.9.7 Norm-2

Norm-2 将每个样本独立地根据其 L2 范数来归一化。

5.9.7.1 输入量

Norm-2 节点输入量如表 87 所示。

表87 Norm-2 节点输入量列表

名称	类型	描述
----	----	----



in_data	DTable	需要进行数据变换的数据
---------	--------	-------------

5.9.7.2 输出量

Norm-2 节点输出量如表 88 所示。

表88 Norm-2 节点输出量列表

名称	类型	描述
out_data	DTable	完成数据变换的数据
model	Model	生成新特征的映射关系模型

5.9.7.3 配置参数



图230 Norm-2 节点属性

Norm-2 节点配置参数如图 230 所示，各参数描述如下：

- 节点名称：节点在建模流程图中的名称标注；
- 节点描述：节点描述信息，用于备注节点所实现的功能；
- 选择算法：Norm-2。

需要注意的是，当该节点未执行时，原始特征量对于该节点是未知的。当该节点被执行时，系统会自动弹出数据变换窗口，用户在弹窗中对新特征生成的规则进行编辑，如图 231 所示。





图231 Norm-2 配置窗口

如图 231 所示，Norm-2 配置弹窗左侧为原始特征名称列表，右侧为用户指定数据变换的规则。

5.9.7.4 使用说明

在数据建模流程中，Norm-2 节点一般位于输出量为 DTable 节点下游，与接收 DTable 作为训练数据的模型训练节点之间。

5.10 自定义提取

5.10.1 表达式

表达式节点是灵活度较高的节点，其基于现有特征（变量），通过计算公式产生新的特征（变量）。

5.10.1.1 输入量

表达式节点输入量如表 89 所示。

表89 表达式节点输入量列表

名称	类型	描述
in_data	DTable	需要进行特征提取的数据

5.10.1.2 输出量

表达式节点输出量如表 90 所示。

表90 表达式节点输出量列表

名称	类型	描述
outdata_data	DTable	增加新特征的数据



5.10.1.3 配置参数



图232 表达式节点属性

表达式节点配置参数如图 232 所示，各参数描述如下：

- 节点名称：节点在建模流程图中的名称标注；
- 节点描述：节点描述信息，用于备注节点所实现的功能；
- 特征变量：新生成特征的名称；
- 表达式：新特征生成时的映射关系。

需要注意的是，当该节点未执行时，原始特征量对于该节点是未知的。当该节点被执行时，系统会自动弹出自定义脚本窗口，用户在弹窗中对新特征生成的映射关系进行编辑，如图 233 所示。



图233 表达式节点配置窗口



如图 233 所示，表达式配置弹窗左侧为自定义特征的名称，右侧为自定义特征的映射关系表达式，其中表达式上方左侧为数据原始特征，上方右侧为基础数学运算符和基础数学函数。

单击原始变量名称或数学函数，其内容即被添加至下方的自定义特征映射关系表达式中。

5.10.1.4 使用说明

在数据建模流程中，表达式节点一般位于输出量为 DTable 节点下游，与接收 DTable 作为训练数据的模型训练节点之间。需要注意的，特征的增加对模型效果影响不是固定的，只有那些对当前场景表征准确的特征才会提升模型的训练效果，因此特征提取往往基于专家知识或工程经验。

5.10.2 数值运算-一元运算

一元运算特征提取节点针对某一个原始特征，支持与生成原始特征成相反数、绝对值关系的新特征。

5.10.2.1 输入量

一元运算节点输入量如表 91 所示。

表91 一元运算节点输入量列表

名称	类型	描述
in_data	DTable	需要进行特征提取的数据

5.10.2.2 输出量

一元运算节点输出量如表 92 所示。

表92 一元运算节点输出量列表

名称	类型	描述
outdata_data	DTable	增加新特征的数据
model	Model	生成新特征的映射关系模型



5.10.2.3 配置参数

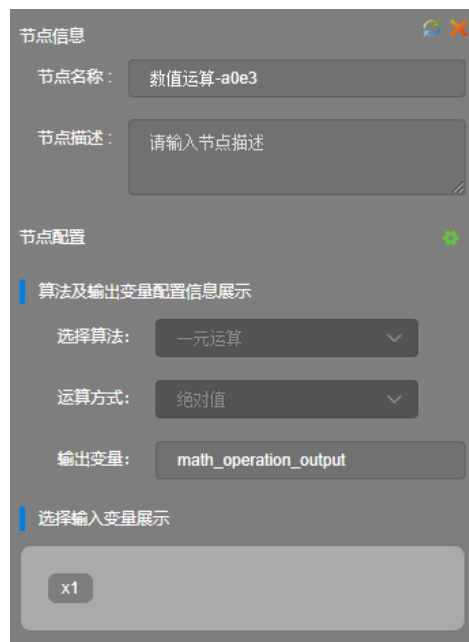


图234 一元运算节点属性

一元运算节点配置参数如图 234 所示，各参数描述如下：

- 节点名称：节点在建模流程图中的名称标注；
- 节点描述：节点描述信息，用于备注节点所实现的功能；
- 选择算法：所选中的运算类型；
- 运算方式：新特征与原始特征之间的关系；
- 输入变量：所选中的某一原始特征（变量）名称；
- 输出变量：基于某一原始特征（变量）生成的新特征（变量）名称。

需要注意的是，当该节点未执行时，原始特征量对于该节点是未知的。当该节点被执行时，系统会自动弹出数值运算窗口，用户在弹窗中对新特征生成的关系进行编辑，如图 235 所示。





图235 一元运算关系

如图 235 所示，一元运算弹窗左侧为原始特征名称列表，右侧列表为被选中的某一个原始特征，右下方为新特征名称及与原始特征的关系。

使用时，首先选中左侧某一原始特征，并将其添加至右侧列表中，配置输出变量名称和运算符后，单击提交按钮，即可实现一元运算特征的提取。

5.10.2.4 使用说明

在数据建模流程中，一元运算节点一般位于输出量为 DTable 节点下游，与接收 DTable 作为训练数据的模型训练节点之间。

5.10.3 数值运算-二元运算

二元运算特征提取节点针对某两个原始特征，支持生成原始特征和、差、绝对差等关系的新特征。

5.10.3.1 输入量

二元运算节点输入量如表 93 所示。

表93 二元运算节点输入量列表

名称	类型	描述
in_data	DTable	需要进行特征提取的数据



5.10.3.2 输出量

二元运算节点输出量如表 94 所示。

表94 二元运算节点输出量列表

名称	类型	描述
outdata_data	DTTable	增加新特征的数据
model	Model	生成新特征的映射关系模型

5.10.3.3 配置参数

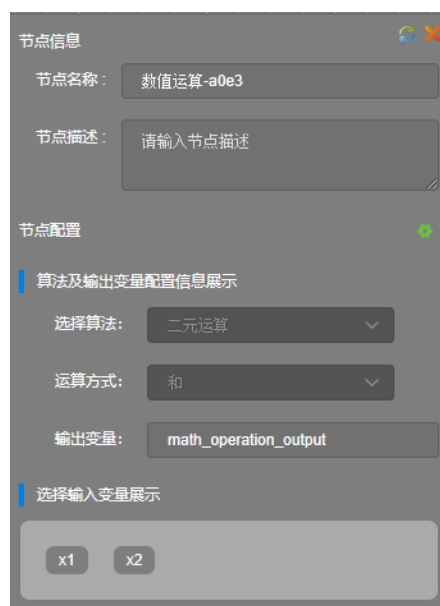


图236 二元节点属性

二元运算节点配置参数如图 236 所示，各参数描述如下：

- 节点名称：节点在建模流程图中的名称标注；
- 节点描述：节点描述信息，用于备注节点所实现的功能；
- 选择算法：所选中的运算类型；
- 运算方式：新特征与原始特征之间的关系；
- 输入变量：所选中的两个原始特征（变量）名称；
- 输出变量：基于两个原始特征（变量）生成的新特征（变量）名称。

需要注意的是，当该节点未执行时，原始特征量对于该节点是未知的。当该节点被执行时，系统会自动弹出二元运算窗口，用户在弹窗中对新特征生成的关系进行编辑，如图 237 所示。





图237 二元运算关系

如图 237 所示，二元运算弹窗左侧为原始特征名称列表，右侧列表为被选中的某两个原始特征，右下方为新特征名称及与原始特征的关系。

使用时，首先选中左侧某两个原始特征，并将其添加至右侧列表中，配置输出变量名称和运算符后，单击提交按钮，即可实现二元运算特征的提取。

5.10.3.4 使用说明

在数据建模流程中，二元运算节点一般位于输出量为 DTable 节点下游，与接收 DTable 作为训练数据的模型训练节点之间。

5.10.4 数值运算-多元运算

多元运算特征提取节点针对多个原始特征，支持生成原始特征最大偏差、平均值等关系的新特征。

5.10.4.1 输入量

多元运算节点输入量如表 95 所示。

表95 多元运算节点输入量列表

名称	类型	描述
in_data	DTable	需要进行特征提取的数据



5.10.4.2 输出量

多元运算节点输出量如表 96 所示。

表96 多元运算节点输出量列表

名称	类型	描述
outdata_data	DTTable	增加新特征的数据
model	Model	生成新特征的映射关系模型

5.10.4.3 配置参数

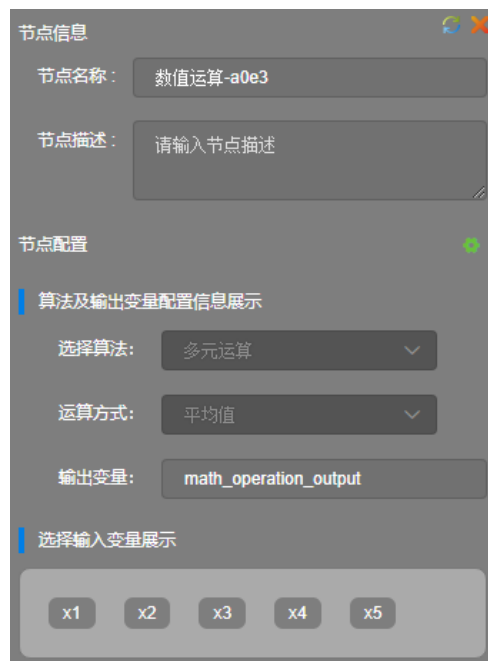


图238 多元运算节点属性

多元运算节点配置参数如图 238 所示，各参数描述如下：

- 节点名称：节点在建模流程图中的名称标注；
- 节点描述：节点描述信息，用于备注节点所实现的功能；
- 选择算法：所选中的运算类型；
- 运算方式：新特征与原始特征之间的关系；
- 输入变量：所选中的多个原始特征（变量）名称；
- 输出变量：基于多个原始特征（变量）生成的新特征（变量）名称。



需要注意的是，当该节点未执行时，原始特征量对于该节点是未知的。当该节点被执行时，系统会自动弹出多元运算窗口，用户在弹窗中对新特征生成的关系进行编辑，如图 239 所示。



图239 多元运算关系

如图 239 所示，多元运算弹窗左侧为原始特征名称列表，右侧列表为被选中的某两个原始特征，右下方为新特征名称及与原始特征的关系。

使用时，首先选中左侧多个原始特征，并将其添加至右侧列表中，配置输出变量名称和运算符后，单击提交按钮，即可实现多元运算特征的提取。

5.10.4.4 使用说明

在数据建模流程中，多元运算节点一般位于输出量为 DTable 节点下游，与接收 DTable 作为训练数据的模型训练节点之间。

5.11 相关性

5.11.1 GRA

GRA 基于灰度矩阵分析，给出变量之间的相关性，相关性越强，相关系数评分越高。

5.11.1.1 输入量

GRA 节点输入量如表 97 所示。

表97 GRA 节点输入量列表



名称	类型	描述
in_data	DTable	需要进行特征选择的数据

5.11.1.2 输出量

GRA 节点输出量如表 98 所示。

表98 GRA 节点输出量列表

名称	类型	描述
out_data	DTable	移除与输出变量相关性低的输入变量后的数据

5.11.1.3 配置参数

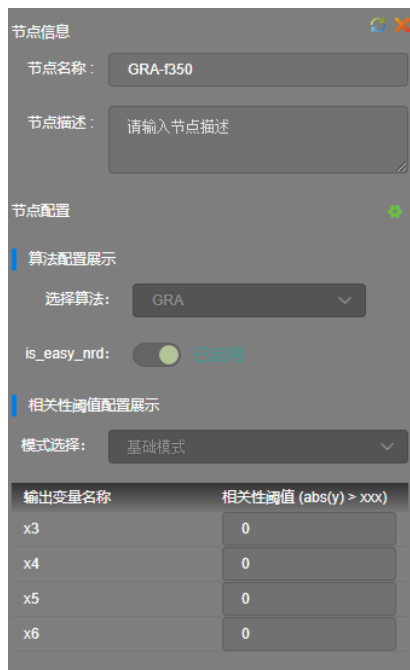


图240 GRA 节点属性

GRA 节点配置参数如图 240 所示，各参数描述如下：

- 节点名称：节点在建模流程图中的名称标注；
- 节点描述：节点描述信息，用于备注节点所实现的功能；
- 算法配置：GRA；
- 节点配置：节点配置分为基础模式和高级模式两种。基础模式根据用户提供输出变量相关性阈值对输入变量进行筛选，如图 240 所示；而高级模式则是根据用户提供的自定义筛选表达式对输入变量进行筛选，需要



注意的是，表示式需基于输出变量，且应满足 python 语法，如图 241 所示。

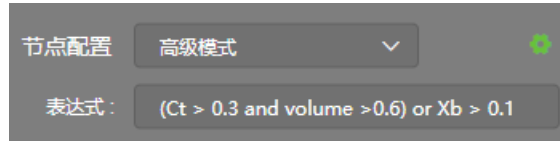


图241 GRA 节点配置高级模式

需要注意的是，当该节点未执行时，变量之间的相关性对于该节点是未知的。当该节点被执行时，系统会自动弹出特征选择窗口，用户在弹窗中对特征（输入变量）进行判定和选择，如图 242 所示。

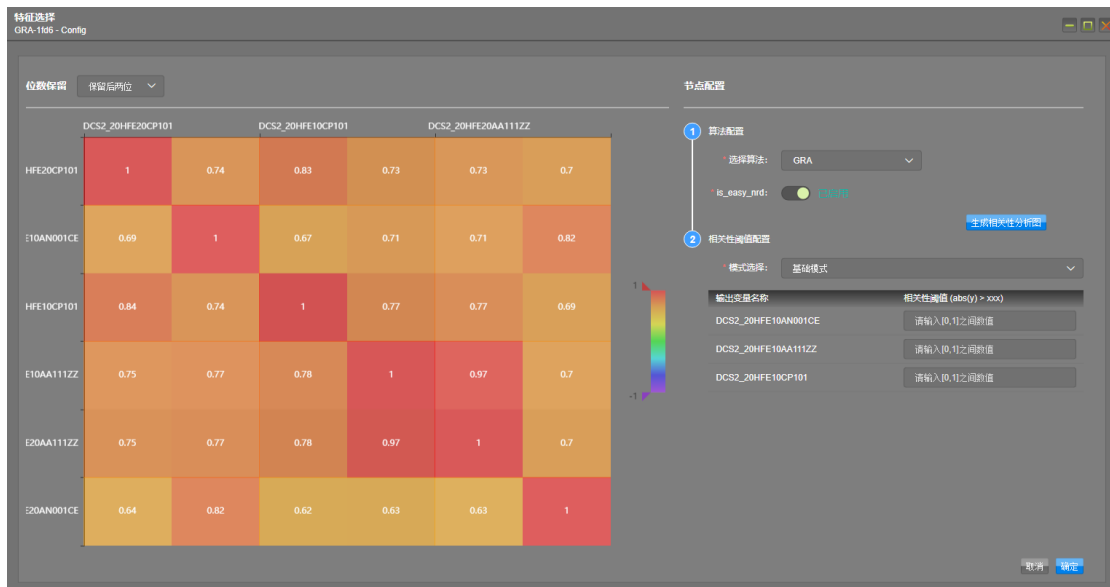


图242 GRA 特征选择窗口

如图 242 所示，特征选择弹窗左侧为由相关性评分绘制得到的热力图，用以辅助用户进行特征选择。右侧为特征选择规则，用户可根据自身需求选择基础模式或高级模式配置选择规则。

5.11.1.4 使用说明

在数据建模流程中，GRA 节点一般位于输出量为 DTable 节点下游，与接收 DTable 作为训练数据的模型训练节点之间。



5.11.2 Correlation

Correlation 基于相关系数分析，给出变量之间的相关性，变量之间从完全负相关到完全正相关相关系数从-1~1 排列。

5.11.2.1 输入量

Correlation 节点输入量如表 99 所示。

表99 Correlation 节点输入量列表

名称	类型	描述
in_data	DTTable	需要进行特征选择的数据

5.11.2.2 输出量

Correlation 节点输出量如表 100 所示。

表100 Correlation 节点输出量列表

名称	类型	描述
out_data	DTTable	移除与输出变量相关性低的输入变量后的数据

5.11.2.3 配置参数

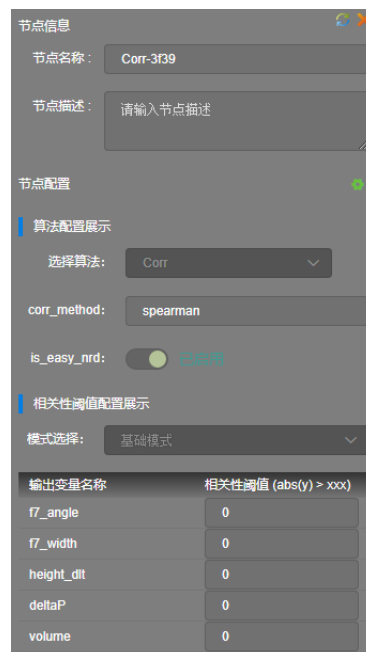


图243 Correlation 节点属性

Correlation 节点配置参数如图 243 所示，各参数描述如下：

- 节点名称：节点在建模流程图中的名称标注；



- 节点描述：节点描述信息，用于备注节点所实现的功能；
- 算法配置：Correlation；
- 节点配置：节点配置分为基础模式和高级模式两种。基础模式根据用户提供输出变量相关性阈值对输入变量进行筛选，如图 243 所示；而高级模式则是根据用户提供的自定义筛选表达式对输入变量进行筛选，需要注意的是，表示式需基于输出变量，且应满足 python 语法，如图 244 所示。



图244 Correlation 高级模式

需要注意的是，当该节点未执行时，变量之间的相关性对于该节点是未知的。当该节点被执行时，系统会自动弹出特征选择窗口，用户在弹窗中对特征（输入变量）进行判定和选择，如图 245 所示。



图245 Correlation 特征选择窗口

如图 245 所示，特征选择弹窗左侧为由相关性评分绘制得到的热力图，用以辅助用户进行特征选择。右侧为特征选择规则，用户可根据自身需求选择基础模式或高级模式配置选择规则。



5.11.2.4 使用说明

在数据建模流程中，Correlation 节点一般位于输出量为 DTable 节点下游，与接收 DTable 作为训练数据的模型训练节点之间。

5.11.3 MIC

MIC 基于互信息分析，给出变量之间的相关性，相关性越强，相关系数评分越高。

5.11.3.1 输入量

MIC 节点输入量如表 101 所示。

表101 MIC 节点输入量列表

名称	类型	描述
in_data	DTable	需要进行特征选择的数据

5.11.3.2 输出量

MIC 节点输出量如表 102 所示。

表102 MIC 节点输出量列表

名称	类型	描述
out_data	DTable	移除与输出变量相关性低的输入变量后的数据



5.11.3.3 配置参数



图246 MIC 节点属性

MIC 节点配置参数如图 246 所示，各参数描述如下：

- 节点名称：节点在建模流程图中的名称标注；
- 节点描述：节点描述信息，用于备注节点所实现的功能；
- 算法配置：MIC；
- 节点配置：节点配置分为基础模式和高级模式两种。基础模式根据用户提供输出变量相关性阈值对输入变量进行筛选，如图 246 所示；而高级模式则是根据用户提供的自定义筛选表达式对输入变量进行筛选，需要注意的是，表示式需基于输出变量，且应满足 python 语法，如图 247 所示。



图247 高级模式



需要注意的是,当该节点未执行时,变量之间的相关性对于该节点是未知的。当该节点被执行时,系统会自动弹出特征选择窗口,用户在弹窗中对特征(输入变量)进行判定和选择,如图 248 所示。



图248 MIC 特征选择窗口

如图 248 所示,特征选择弹窗左侧为由相关性评分绘制得到的热力图,用以辅助用户进行特征选择。右侧为特征选择规则,用户可根据自身需求选择基础模式或高级模式配置选择规则。

5.11.3.4 使用说明

在数据建模流程中,MIC 节点一般位于输出量为 DTable 节点下游,与接收 DTable 作为训练数据的模型训练节点之间。

5.12 重要性

5.12.1 MDA

MDA 通过回归方法拟合数据,以此得到数据集中各个特征的系数/信息增益比例,从而得到特征的重要性排序。根据回归方法的不同,特征的排序情况也会不同。与其他特征重要性评估的区别在于,MDA 使用拟合系数的大小进行排序。

5.12.1.1 输入量

MDA 节点输入量如表 103 所示。

表103 MDA 节点输入量列表



名称	类型	描述
in_data	DTTable	需要进行特征重要性评估的数据

5.12.1.2 输出量

MDA 节点输出量如表 104 所示。

表104 MDA 节点输出量列表

名称	类型	描述
out_data	DTTable	移除与输出变量重要性低的输入变量后的数据

5.12.1.3 配置参数



图249 MDA 节点属性

MDA 节点配置参数如图 249 所示，各参数描述如下：

- 节点名称：节点在建模流程图中的名称标注；
- 节点描述：节点描述信息，用于备注节点所实现的功能；
- 算法配置：MDA；
- func_name：使用的回归方法，在当前版本固定为 adaboost；
- 随机种子：可以为空或输入正整数，后者用于指定算法超参数初始值生成的规则；



- 重要性阈值：输入大于 0 的数值，小于该重要性阈值的输入变量会在节点输出结果中被剔除。

需要注意的是，当该节点未执行时，变量之间的重要性对于该节点是未知的。当该节点被执行时，系统会自动弹出特征选择窗口，用户在弹窗中对特征（输入变量）进行判定和选择，如图 250 所示。各功能按钮描述如下：

- 生成重要性分析图：根据选择的特征重要性算法及回归算法，生成左侧重要性分析图（条形图）；
- 位数保留：切换在重要性分析图中，鼠标在对应变量悬停时显示的重要性数值精度；
- 排序：切换在重要性分析图中，变量的先后顺序是基于输入变量顺序或基于重要性数值顺序。

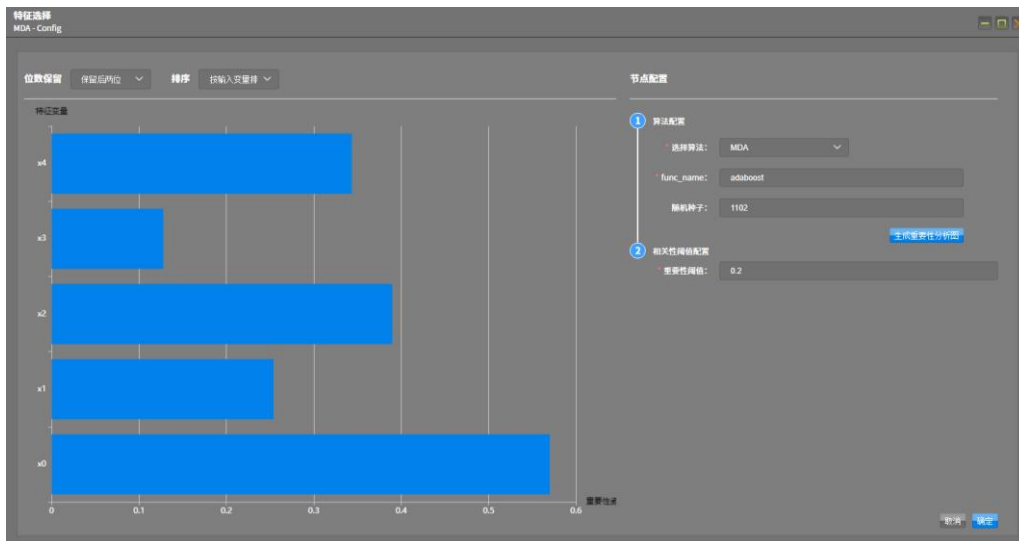


图250 MDA 配置窗口

5.12.1.4 使用说明

在数据建模流程中，MDA 节点一般位于输出量为 DTable 节点下游，与接收 DTable 作为训练数据的模型训练节点之间。

5.12.2 MDI

MDI 通过回归方法拟合数据，以此得到数据集中各个特征的系数/信息增益比例，从而得到特征的重要性排序。根据回归方法的不同，特征的排序情况也会



不同。与其他特征重要性评估的区别在于，MDI 通过每次加入的特征所带来的信息增益进行排序。

5.12.2.1 输入量

MDI 节点输入量如表 105 所示。

表105 MDI 节点输入量列表

名称	类型	描述
in_data	DTTable	需要进行特征重要性评估的数据

5.12.2.2 输出量

MDI 节点输出量如表 106 所示。

表106 MDI 节点输出量列表

名称	类型	描述
out_data	DTTable	移除与输出变量重要性低的输入变量后的数据

5.12.2.3 配置参数



图251 MDI 节点属性

MDI 节点配置参数如图 251 所示，各参数描述如下：

- 节点名称：节点在建模流程图中的名称标注；



- 节点描述：节点描述信息，用于备注节点所实现的功能；
- 算法配置：MDI；
- func_name：使用的回归方法，在当前版本固定为 adaboost；
- 随机种子：可以为空或输入正整数，后者用于指定算法超参数初始值生成的规则；
- 重要性阈值：输入大于 0 的数值，小于该重要性阈值的输入变量会在节点输出结果中被剔除。

需要注意的是，当该节点未执行时，变量之间的重要性对于该节点是未知的。当该节点被执行时，系统会自动弹出特征选择窗口，用户在弹窗中对特征（输入变量）进行判定和选择，如图 252 所示。各功能按钮描述如下：

- 生成重要性分析图：根据选择的特征重要性算法及回归算法，生成左侧重要性分析图（条形图）；
- 位数保留：切换在重要性分析图中，鼠标在对应变量悬停时显示的重要性数值精度；
- 排序：切换在重要性分析图中，变量的先后顺序是基于输入变量顺序或基于重要性数值顺序。

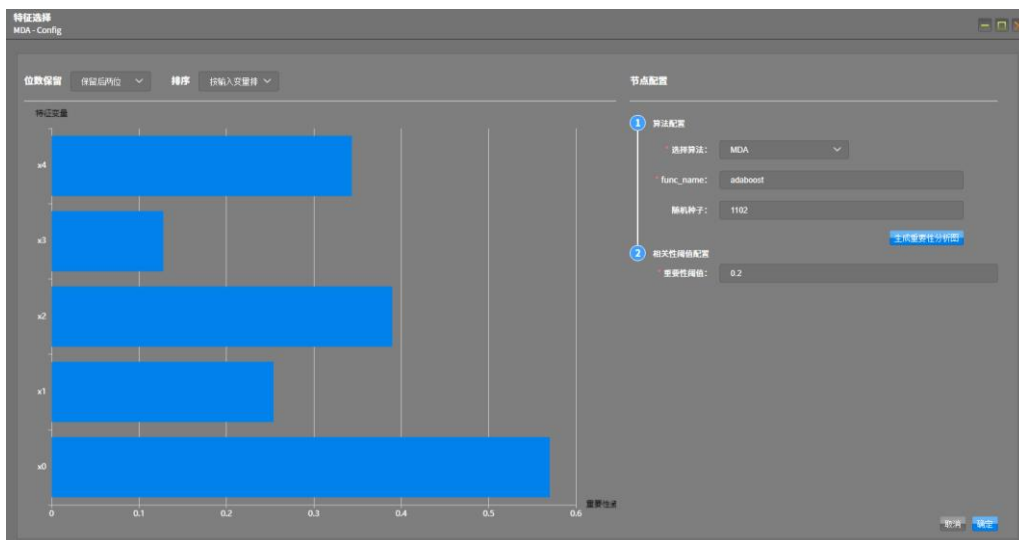


图252 MDI 配置窗口



5.12.2.4 使用说明

在数据建模流程中，MDI 节点一般位于输出量为 DTable 节点下游，与接收 DTable 作为训练数据的模型训练节点之间。

5.13 AI Agent

5.13.1 AI Agent

依托 TF-AutoML 技术，DT-Empower 内嵌智能模型训练算法 AI Agent，该算法将模型选择、超参选择、数据训练等各个过程进行组合优化，向用户提供全自动、零门槛的模型训练功能。此外强化学习技术的集成加强了算法空间搜索的目标性，大幅缩短了模型迭代训练时间。

5.13.1.1 输入量

AI Agent 节点输入量如表 107 所示。

表107 AI Agent 节点输入量列表

名称	类型	描述
in_data	DTable	带有输入输出变量标识的数据

5.13.1.2 输出量

AI Agent 节点输出量如表 108 所示。

表108 AI Agent 节点输出量列表

名称	类型	描述
predictions	DTable	输出变量真实值、模型预测值及预测误差数据
model	Model	训练后得到的模型
meta_model	MetaModel	模型的超参数



5.13.1.3 配置参数



图253 AI Agent 节点属性

AI Agent 节点配置参数如图 253 所示，各参数描述如下：

- 节点名称：节点在建模流程图中的名称标注；
- 节点描述：节点描述信息，用于备注节点所实现的功能；
- 超参数优化：超参数自优化功能，该功能开启后带来模型细节精度提升的同时，会在一定程度上增加训练耗时。

5.13.1.4 使用说明

在数据建模流程中，AI Agent 节点一般位于数据清洗和特征选择类节点的下流，模型（知识）聚合节点的上游。经 AI Agent 训练得到模型也可以独立存储应用。

5.14 线性拟合

5.14.1 BayesianRidge

BayesianRidge 节点基于线性关系进行模型训练。

5.14.1.1 输入量

BayesianRidge 节点输入量如表 109 所示。

表109 BayesianRidge 节点输入量列表

名称	类型	描述
in_data	DTTable	带有输入输出变量标识的数据



5.14.1.2 输出量

BayesianRidge 节点输出量如表 110 所示。

表110 BayesianRidge 节点输出量列表

名称	类型	描述
predictions	DTTable	输出变量真实值、模型预测值及预测误差数据
model	Model	训练后得到的模型
meta_model	MetaModel	模型的超参数

5.14.1.3 配置参数



图254 BayesianRidge 节点属性

BayesianRidge 节点配置参数如图 254 所示，各参数描述如下：

- 节点名称：节点在建模流程图中的名称标注；
- 节点描述：节点描述信息，用于备注节点所实现的功能；
- 基础配置：指定模型训练算法中的基础参数配置；
- 超参优化：开启超参优化并进入高级配置；
- 高级配置：指定模型训练算法中的高级参数配置。



5.14.1.4 使用说明

在数据建模流程中，BayesianRidge 节点一般位于数据清洗和特征选择类节点的下游，模型（知识）聚合节点的上游。经线性拟合训练得到模型也可以独立存储应用。

需要注意的是 BayesianRidge 线性拟合模型仅在输入变量与输出变量呈线性关系时效果才最好。

5.14.2 ElasticNet

ElasticNet 节点基于线性关系进行模型训练。

5.14.2.1 输入量

ElasticNet 节点输入量如表 111 所示。

表111 ElasticNet 节点输入量列表

名称	类型	描述
in_data	DTTable	带有输入输出变量标识的数据

5.14.2.2 输出量

ElasticNet 节点输出量如表 112 所示。

表112 ElasticNet 节点输出量列表

名称	类型	描述
predictions	DTTable	输出变量真实值、模型预测值及预测误差数据
model	Model	训练后得到的模型
meta_model	MetaModel	模型的超参数



5.14.2.3 配置参数



图255 ElasticNet 节点属性

ElasticNet 节点配置参数如图 255 所示，各参数描述如下：

- 节点名称：节点在建模流程图中的名称标注；
- 节点描述：节点描述信息，用于备注节点所实现的功能；
- 基础配置：指定模型训练算法中的基础参数配置；
- 超参优化：开启超参优化并进入高级配置；
- 高级配置：指定模型训练算法中的高级参数配置。

5.14.2.4 使用说明

在数据建模流程中，ElasticNet 节点一般位于数据清洗和特征选择类节点的下游，模型（知识）聚合节点的上游。经线性拟合训练得到模型也可以独立存储应用。

需要注意的是 ElasticNet 线性拟合模型仅在输入变量与输出变量呈线性关系时效果才最好。

5.14.3 ElasticNetCV

ElasticNetCV 节点基于线性关系进行模型训练。



5.14.3.1 输入量

ElasticNetCV 节点输入量如表 113 所示。

表113 ElasticNetCV 节点输入量列表

名称	类型	描述
in_data	DTTable	带有输入输出变量标识的数据

5.14.3.2 输出量

ElasticNetCV 节点输出量如表 114 所示。

表114 ElasticNetCV 节点输出量列表

名称	类型	描述
predictions	DTTable	输出变量真实值、模型预测值及预测误差数据
model	Model	训练后得到的模型
meta_model	MetaModel	模型的超参数

5.14.3.3 配置参数



图256 ElasticNetCV 节点属性

ElasticNetCV 节点配置参数如图 256 所示，各参数描述如下：

- 节点名称：节点在建模流程图中的名称标注；



- 节点描述：节点描述信息，用于备注节点所实现的功能；
- 基础配置：指定模型训练算法中的基础参数配置；
- 超参优化：开启超参优化并进入高级配置；
- 高级配置：指定模型训练算法中的高级参数配置。

5.14.3.4 使用说明

在数据建模流程中，ElasticNetCV 节点一般位于数据清洗和特征选择类节点的下流，模型（知识）聚合节点的上游。经线性拟合训练得到模型也可以独立存储应用。

需要注意的是 ElasticNetCV 线性拟合模型仅在输入变量与输出变量呈线性关系时效果才最好。

5.14.4 Huber

Huber 节点基于线性关系进行模型训练。

5.14.4.1 输入量

Huber 节点输入量如表 115 所示。

表115 Huber 节点输入量列表

名称	类型	描述
in_data	DTTable	带有输入输出变量标识的数据

5.14.4.2 输出量

Huber 节点输出量如表 116 所示。

表116 Huber 节点输出量列表

名称	类型	描述
predictions	DTTable	输出变量真实值、模型预测值及预测误差数据
model	Model	训练后得到的模型
meta_model	MetaModel	模型的超参数



5.14.4.3 配置参数

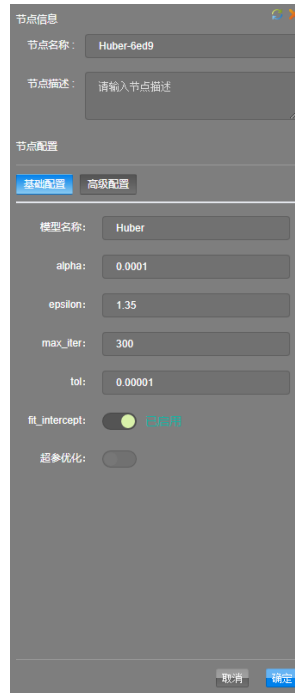


图257 Huber 节点属性

Huber 节点配置参数如图 257 所示，各参数描述如下：

- 节点名称：节点在建模流程图中的名称标注；
- 节点描述：节点描述信息，用于备注节点所实现的功能；
- 基础配置：指定模型训练算法中的基础参数配置；
- 超参优化：开启超参优化并进入高级配置；
- 高级配置：指定模型训练算法中的高级参数配置。

5.14.4.4 使用说明

在数据建模流程中，Huber 节点一般位于数据清洗和特征选择类节点的下游，模型（知识）聚合节点的上游。经线性拟合训练得到模型也可以独立存储应用。

需要注意的是 Huber 线性拟合模型仅在输入变量与输出变量呈线性关系时效果才最好。

5.14.5 LAR

LAR 节点基于线性关系进行模型训练。



5.14.5.1 输入量

LAR 节点输入量如表 117 所示。

表117 LAR 节点输入量列表

名称	类型	描述
in_data	DTTable	带有输入输出变量标识的数据

5.14.5.2 输出量

LAR 节点输出量如表 118 所示。

表118 LAR 节点输出量列表

名称	类型	描述
predictions	DTTable	输出变量真实值、模型预测值及预测误差数据
model	Model	训练后得到的模型
meta_model	MetaModel	模型的超参数

5.14.5.3 配置参数



图258 LAR 节点属性

LAR 节点配置参数如图 258 所示，各参数描述如下：

- 节点名称：节点在建模流程图中的名称标注；



- 节点描述：节点描述信息，用于备注节点所实现的功能；
- 基础配置：指定模型训练算法中的基础参数配置；
- 超参优化：开启超参优化并进入高级配置；
- 高级配置：指定模型训练算法中的高级参数配置。

5.14.5.4 使用说明

在数据建模流程中，LAR 节点一般位于数据清洗和特征选择类节点的下游，模型（知识）聚合节点的上游。经线性拟合训练得到模型也可以独立存储应用。

需要注意的是 LAR 线性拟合模型仅在输入变量与输出变量呈线性关系时效果才最好。

5.14.6 Lasso

Lasso 节点基于线性关系进行模型训练。

5.14.6.1 输入量

Lasso 节点输入量如表 119 所示。

表119 Lasso 节点输入量列表

名称	类型	描述
in_data	DTTable	带有输入输出变量标识的数据

5.14.6.2 输出量

Lasso 节点输出量如表 120 所示。

表120 Lasso 节点输出量列表

名称	类型	描述
predictions	DTTable	输出变量真实值、模型预测值及预测误差数据
model	Model	训练后得到的模型
meta_model	MetaModel	模型的超参数



5.14.6.3 配置参数

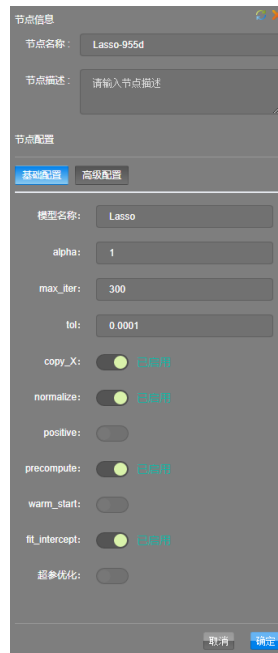


图259 Lasso 节点属性

Lasso 节点配置参数如图 259 所示，各参数描述如下：

- 节点名称：节点在建模流程图中的名称标注；
- 节点描述：节点描述信息，用于备注节点所实现的功能；
- 基础配置：指定模型训练算法中的基础参数配置；
- 超参优化：开启超参优化并进入高级配置；
- 高级配置：指定模型训练算法中的高级参数配置。

5.14.6.4 使用说明

在数据建模流程中，Lasso 节点一般位于数据清洗和特征选择类节点的下游，模型（知识）聚合节点的上游。经线性拟合训练得到模型也可以独立存储应用。

需要注意的是 Lasso 线性拟合模型仅在输入变量与输出变量呈线性关系时效果才最好。

5.14.7 LassoLar

LassoLar 节点基于线性关系进行模型训练。

5.14.7.1 输入量

LassoLar 节点输入量如表 121 所示。



表121 LassoLar 节点输入量列表

名称	类型	描述
in_data	DTTable	带有输入输出变量标识的数据

5.14.7.2 输出量

LassoLar 节点输出量如表 122 所示。

表122 LassoLar 节点输出量列表

名称	类型	描述
predictions	DTTable	输出变量真实值、模型预测值及预测误差数据
model	Model	训练后得到的模型
meta_model	MetaModel	模型的超参数

5.14.7.3 配置参数

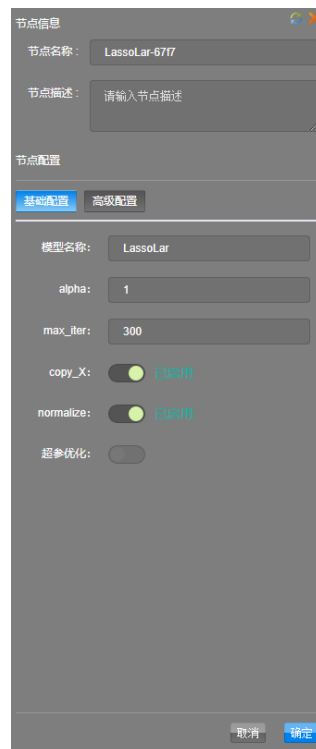


图260 LassoLar 节点属性

LassoLar 节点配置参数如图 260 所示，各参数描述如下：

- 节点名称：节点在建模流程图中的名称标注；
- 节点描述：节点描述信息，用于备注节点所实现的功能；
- 基础配置：指定模型训练算法中的基础参数配置；



- 超参优化：开启超参优化并进入高级配置；
- 高级配置：指定模型训练算法中的高级参数配置。

5.14.7.4 使用说明

在数据建模流程中，LassoLar 节点一般位于数据清洗和特征选择类节点的下游，模型（知识）聚合节点的上游。经线性拟合训练得到模型也可以独立存储应用。

需要注意的是 LassoLar 线性拟合模型仅在输入变量与输出变量呈线性关系时效果才最好。

5.14.8ARD

ARD 节点基于线性关系进行模型训练。

5.14.8.1 输入量

ARD 节点输入量如表 123 所示。

表123 ARD 节点输入量列表

名称	类型	描述
in_data	DTTable	带有输入输出变量标识的数据

5.14.8.2 输出量

ARD 节点输出量如表 124 所示。

表124 ARD 节点输出量列表

名称	类型	描述
predictions	DTTable	输出变量真实值、模型预测值及预测误差数据
model	Model	训练后得到的模型
meta_model	MetaModel	模型的超参数



5.14.8.3 配置参数



图261 ARD 节点属性

ARD 节点配置参数如图 261 所示，各参数描述如下：

- 节点名称：节点在建模流程图中的名称标注；
- 节点描述：节点描述信息，用于备注节点所实现的功能；
- 基础配置：指定模型训练算法中的基础参数配置；
- 超参优化：开启超参优化并进入高级配置；
- 高级配置：指定模型训练算法中的高级参数配置。

5.14.8.4 使用说明

在数据建模流程中，ARD 节点一般位于数据清洗和特征选择类节点的下游，模型（知识）聚合节点的上游。经线性拟合训练得到模型也可以独立存储应用。

需要注意的是 ARD 线性拟合模型仅在输入变量与输出变量呈线性关系时效果才最好。

5.14.9 Linear

Linear 节点基于线性关系进行模型训练。

5.14.9.1 输入量

Linear 节点输入量如表 125 所示。



表125 Linear 节点输入量列表

名称	类型	描述
in_data	DTTable	带有输入输出变量标识的数据

5.14.9.2 输出量

Linear 节点输出量如表 126 所示。

表126 Linear 节点输出量列表

名称	类型	描述
predictions	DTTable	输出变量真实值、模型预测值及预测误差数据
model	Model	训练后得到的模型
meta_model	MetaModel	模型的超参数

5.14.9.3 配置参数



图262 Linear 节点属性

Linear 节点配置参数如图 262 所示，各参数描述如下：

- 节点名称：节点在建模流程图中的名称标注；
- 节点描述：节点描述信息，用于备注节点所实现的功能；



- 基础配置：指定模型训练算法中的基础参数配置；
- 超参优化：开启超参优化并进入高级配置；
- 高级配置：指定模型训练算法中的高级参数配置。

5.14.9.4 使用说明

在数据建模流程中，Linear 节点一般位于数据清洗和特征选择类节点的下流，模型（知识）聚合节点的上游。经线性拟合训练得到模型也可以独立存储应用。

需要注意的是 Linear 线性拟合模型仅在输入变量与输出变量呈线性关系时效果才最好。

5.14.10 Ridge

Ridge 节点基于线性关系进行模型训练。

5.14.10.1 输入量

Ridge 节点输入量如表 127 所示。

表127 Ridge 节点输入量列表

名称	类型	描述
in_data	DTTable	带有输入输出变量标识的数据

5.14.10.2 输出量

Ridge 节点输出量如表 128 所示。

表128 Ridge 节点输出量列表

名称	类型	描述
predictions	DTTable	输出变量真实值、模型预测值及预测误差数据
model	Model	训练后得到的模型
meta_model	MetaModel	模型的超参数



5.14.10.3 配置参数

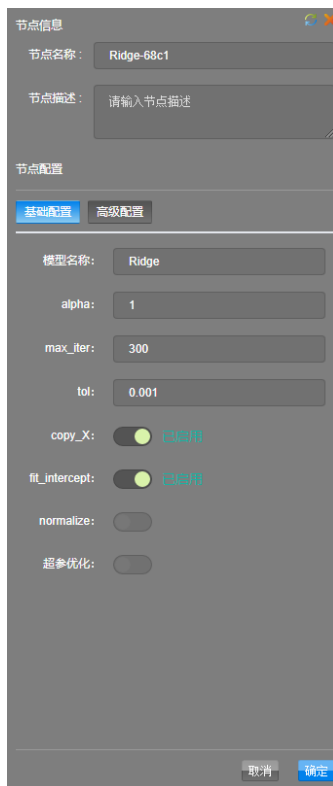


图263 Ridge 节点属性

Ridge 节点配置参数如图 263 所示，各参数描述如下：

- 节点名称：节点在建模流程图中的名称标注；
- 节点描述：节点描述信息，用于备注节点所实现的功能；
- 基础配置：指定模型训练算法中的基础参数配置；
- 超参优化：开启超参优化并进入高级配置；
- 高级配置：指定模型训练算法中的高级参数配置。

5.14.10.4 使用说明

在数据建模流程中，Ridge 节点一般位于数据清洗和特征选择类节点的下流，模型（知识）聚合节点的上游。经线性拟合训练得到模型也可以独立存储应用。

需要注意的是 Ridge 线性拟合模型仅在输入变量与输出变量呈线性关系时效果才最好。

5.14.11 SGD

SGD 节点基于线性关系进行模型训练。



5.14.11.1 输入量

SGD 节点输入量如表 129 所示。

表129 SGD 节点输入量列表

名称	类型	描述
in_data	DTTable	带有输入输出变量标识的数据

5.14.11.2 输出量

SGD 节点输出量如表 130 所示。

表130 SGD 节点输出量列表

名称	类型	描述
predictions	DTTable	输出变量真实值、模型预测值及预测误差数据
model	Model	训练后得到的模型
meta_model	MetaModel	模型的超参数

5.14.11.3 配置参数



图264 SGD 节点属性



SGD 节点配置参数如图 264 所示，各参数描述如下：

- 节点名称：节点在建模流程图中的名称标注；
- 节点描述：节点描述信息，用于备注节点所实现的功能；
- 基础配置：指定模型训练算法中的基础参数配置；
- 超参优化：开启超参优化并进入高级配置；
- 高级配置：指定模型训练算法中的高级参数配置。

5.14.11.4 使用说明

在数据建模流程中，SGD 节点一般位于数据清洗和特征选择类节点的下游，模型（知识）聚合节点的上游。经线性拟合训练得到模型也可以独立存储应用。

需要注意的是 SGD 线性拟合模型仅在输入变量与输出变量呈线性关系时效果才最好。

5.15 非线性拟合

5.15.1 AdaBoost

基于多个回归模型的 Boosting 集成学习方法。这是一种通过样本误差和学习器误差，更新下一个学习器的权重的串行集成学习算法。依次构造弱学习器，通过每个弱学习器中加权误差得到该学习器在模型集成时的权重；并通过前一个学习器中每个样本的误差来更新下一个学习器的样本权重。

5.15.1.1 输入量

AdaBoost 节点输入量如表 131 所示。

表131 AdaBoost 节点输入量列表

名称	类型	描述
in_data	DTTable	带有输入输出变量标识的数据

5.15.1.2 输出量

AdaBoost 节点输出量如表 132 所示。

表132 AdaBoost 节点输出量列表

名称	类型	描述
----	----	----



predictions	DTable	输出变量真实值、模型预测值及预测误差数据
model	Model	训练后得到的模型
meta_model	MetaModel	模型的超参数

5.15.1.3 配置参数

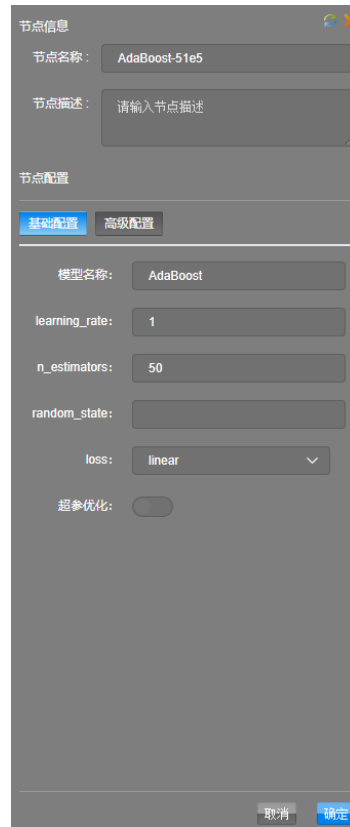


图265 AdaBoost 节点属性

AdaBoost 节点配置参数如图 265 所示，各参数描述如下：

- 节点名称：节点在建模流程图中的名称标注；
 - 节点描述：节点描述信息，用于备注节点所实现的功能；
 - 基础配置：指定模型训练算法中的基础参数配置；
 - 超参优化：开启超参优化并进入超参优化 Tab 页；
 - 交叉验证：指定是否在训练中开启交叉验证；开启后，将在模型详情页显示定性精度指标；
 - 高级配置：指定模型训练算法中的高级参数配置。
- 基础配置中的算法参数介绍如下：



- `n_estimators`: 基学习器的最大数目;
- `learning_rate`: 基学习器的学习率, 与 `n_estimators` 的大小相互制衡;
- `loss`: 基学习器更新权重时的误差计算函数。

5.15.1.4 使用说明

在数据建模流程中, `AdaBoost` 节点一般位于数据清洗和特征选择类节点的下流, 模型(知识)聚合节点的上游。`AdaBoost` 训练得到模型也可以独立存储应用。

5.15.2 BaggingDeFo

基于多个回归模型的 `Bagging` 集成学习方法。这是一种通过对样本和特征的随机抽样, 以此构造若干个弱学习器, 从而提高模型在数据集上的泛化能力的并行集成学习算法。通过对训练集样本和外源性特征分别进行抽样, 根据抽样后的数据建立弱学习器, 重复若干次以上步骤, 最后将多个学习器的结果按照权重(一般为相同权重)进行组合。

5.15.2.1 输入量

`BaggingDeFo` 节点输入量如表 133 所示。

表133 BaggingDeFo 节点输入量列表

名称	类型	描述
<code>in_data</code>	<code>DTTable</code>	带有输入输出变量标识的数据

5.15.2.2 输出量

`BaggingDeFo` 节点输出量如表 134 所示。

表134 BaggingDeFo 节点输出量列表

名称	类型	描述
<code>predictions</code>	<code>DTTable</code>	输出变量真实值、模型预测值及预测误差数据
<code>model</code>	<code>Model</code>	训练后得到的模型
<code>meta_model</code>	<code>MetaModel</code>	模型的超参数



5.15.2.3 配置参数



图266 BaggingDeFo 节点属性

BaggingDeFo 节点配置参数如图 266 所示，各参数描述如下：

- 节点名称：节点在建模流程图中的名称标注；
- 节点描述：节点描述信息，用于备注节点所实现的功能；
- 基础配置：指定模型训练算法中的基础参数配置；
- 超参优化：开启超参优化并进入超参优化 Tab 页；
- 交叉验证：指定是否在训练中开启交叉验证；开启后，将在模型详情页显示定性精度指标；

- 高级配置：指定模型训练算法中的高级参数配置。

基础配置中的算法参数介绍如下：

- n_estimators：基学习器的最大数目；
- max_samples：每个基学习器使用的样本数量或者比例；
- max_features：每个基学习器使用的特征数量或者比例；



- **bootstrap**: 是否对样本进行有放回的抽样, 若为 **False**, 进行无放回的抽样;
- **bootstrap_features**: 是否对特征集合进行有放回的抽样;
- **oob_score**: 误差计算方式, 是否使用袋外样本计算误差。

5.15.2.4 使用说明

在数据建模流程中, **BaggingDeFo** 节点一般位于数据清洗和特征选择类节点的下游, 模型(知识)聚合节点的上游。**BaggingDeFo** 训练得到模型也可以独立存储应用。

5.15.3 CART

CART 是一种二叉树模型, 它通过平方误差选择最优特征和最优特征下最佳的切分点, 并以叶节点中样本的均值作为预测输出。

5.15.3.1 输入量

CART 节点输入量如表 135 所示。

表135 CART 节点输入量列表

名称	类型	描述
in_data	DTTable	带有输入输出变量标识的数据

5.15.3.2 输出量

CART 节点输出量如表 136 所示。

表136 CART 节点输出量列表

名称	类型	描述
predictions	DTTable	输出变量真实值、模型预测值及预测误差数据
model	Model	训练后得到的模型
meta_model	MetaModel	模型的超参数



5.15.3.3 配置参数



图267 CART 节点属性

CART 节点配置参数如图 267 所示，各参数描述如下：

- 节点名称：节点在建模流程图中的名称标注；
- 节点描述：节点描述信息，用于备注节点所实现的功能；
- 基础配置：指定模型训练算法中的基础参数配置；
- 超参优化：开启超参优化并进入超参优化 Tab 页；
- 交叉验证：指定是否在训练中开启交叉验证；开启后，将在模型详情页

显示定性精度指标；

- 高级配置：指定模型训练算法中的高级参数配置。

基础配置中的算法参数介绍如下：

- min_samples_split：拆分内部节点所需的最少样本数；
- min_samples_leaf：一个叶节点中的最小样本个数；
- max_features：节点分裂时参与判断的最大特征数；
- min_impurity_decrease：设定每次分支基尼系数增大/误差减少（回归问题）

的阈值，不超过这个阈值的分支不会进行。



5.15.3.4 使用说明

在数据建模流程中，CART 节点一般位于数据清洗和特征选择类节点的下游，模型（知识）聚合节点的上游。CART 训练得到模型也可以独立存储应用。

5.15.4 ExtraTrees

极端回归树。这是一种随机选择特征进行分支的集成学习算法，相比于 Bagging，ExtraTrees 具有更强的泛化能力。通过随机选择的方法选择每一个节点分支特征，以提高算法的泛化能力；但与此同时，单个基学习器的精度比上述两个集成算法更低，所以 ExtraTrees 需要更多的基学习器。

5.15.4.1 输入量

ExtraTrees 节点输入量如表 137 所示。

表137 ExtraTrees 节点输入量列表

名称	类型	描述
in_data	DTTable	带有输入输出变量标识的数据

5.15.4.2 输出量

ExtraTrees 节点输出量如表 138 所示。

表138 ExtraTrees 节点输出量列表

名称	类型	描述
predictions	DTTable	输出变量真实值、模型预测值及预测误差数据
model	Model	训练后得到的模型
meta_model	MetaModel	模型的超参数



5.15.4.3 配置参数



图268 ExtraTrees 节点属性

ExtraTrees 节点配置参数如图 268 所示，各参数描述如下：

- 节点名称：节点在建模流程图中的名称标注；
- 节点描述：节点描述信息，用于备注节点所实现的功能；
- 基础配置：指定模型训练算法中的基础参数配置；
- 超参优化：开启超参优化并进入超参优化 Tab 页；
- 交叉验证：指定是否在训练中开启交叉验证；开启后，将在模型详情页显示定性精度指标；

- 高级配置：指定模型训练算法中的高级参数配置。

基础配置中的算法参数介绍如下：

- `n_estimators`：基学习器的最大数目；
- `criterion`：测量节点分割质量的函数；
- `bootstrap`：是否对所有样本进行有放回的抽样，若为 `False`，使用全部样本进行训练；
- `oob_score`：误差计算方式，是否使用袋外的样本计算 `R2` 指标。



5.15.4.4 使用说明

在数据建模流程中，ExtraTrees 节点一般位于数据清洗和特征选择类节点的下流，模型（知识）聚合节点的上游。ExtraTrees 训练得到模型也可以独立存储应用。

5.15.5 GP

高斯过程回归（Gaussian Process Regression, GPR）是使用高斯过程先验对数据进行回归分析的非参数模型。高斯过程回归通过核函数来定义协方差函数，再通过协方差函数构建协方差矩阵，得到输入变量与输出变量之间的联合分布，最后通过这个联合分布计算未知样本对应的输出变量。

5.15.5.1 输入量

GP 节点输入量如表 139 所示。

表139 GP 节点输入量列表

名称	类型	描述
in_data	DTTable	带有输入输出变量标识的数据

5.15.5.2 输出量

GP 节点输出量如表 140 所示。

表140 GP 节点输出量列表

名称	类型	描述
predictions	DTTable	输出变量真实值、模型预测值及预测误差数据
model	Model	训练后得到的模型
meta_model	MetaModel	模型的超参数



5.15.5.3 配置参数



图269 GP 节点属性

GP 节点配置参数如图 269 所示，各参数描述如下：

- 节点名称：节点在建模流程图中的名称标注；
- 节点描述：节点描述信息，用于备注节点所实现的功能；
- 基础配置：指定模型训练算法中的基础参数配置；
- 超参优化：开启超参优化并进入超参优化 Tab 页；
- 交叉验证：指定是否在训练中开启交叉验证；开启后，将在模型详情页显示定性精度指标；

- 高级配置：指定模型训练算法中的高级参数配置。

基础配置中的算法参数介绍如下：

- **alpha**：拟合期间添加到核矩阵对角线上的值，并且可以确保计算值形成正定矩阵，防止拟合过程中出现潜在的数值问题；
- **n_restarts_optimizer**：优化器重新启动的次数，用于寻找最大化对数边际可能性的内核参数。



5.15.5.4 使用说明

在数据建模流程中，GP 节点一般位于数据清洗和特征选择类节点的下游，模型（知识）聚合节点的上游。GP 训练得到模型也可以独立存储应用。

5.15.6 GBDT

梯度回归树。这是一种通过学习前一次拟合结果的残差，逐渐优化预测精度的串行集成学习算法。依次构造弱学习器，将前 k-1 个学习器拟合结果的残差作为第 k 个学习器的输入，从而逐步提升预测的精度，可以有效减小预测的误差

5.15.6.1 输入量

GBDT 节点输入量如表 141 所示。

表141 GBDT 节点输入量列表

名称	类型	描述
in_data	DTTable	带有输入输出变量标识的数据

5.15.6.2 输出量

GBDT 节点输出量如表 142 所示。

表142 GBDT 节点输出量列表

名称	类型	描述
predictions	DTTable	输出变量真实值、模型预测值及预测误差数据
model	Model	训练后得到的模型
meta_model	MetaModel	模型的超参数



5.15.6.3 配置参数



图270 GBDT 节点属性

GBDT 节点配置参数如图 270 所示，各参数描述如下：

- 节点名称：节点在建模流程图中的名称标注；
- 节点描述：节点描述信息，用于备注节点所实现的功能；
- 基础配置：指定模型训练算法中的基础参数配置；
- 超参优化：开启超参优化并进入超参优化 Tab 页；
- 交叉验证：指定是否在训练中开启交叉验证；开启后，将在模型详情页显示定性精度指标；

- 高级配置：指定模型训练算法中的高级参数配置。

基础配置中的算法参数介绍如下：

- loss：在提升过程中使用的损失函数；
- learning_rate：基学习器的学习率，与 n_estimators 的大小相互制衡；
- n_estimators：基学习器的最大数目，亦是 Boosting 过程的最大迭代次数；
- subsample：基学习器中使用的样本比例；



- **criterion**: 测量节点分割质量的函数。

5.15.6.4 使用说明

在数据建模流程中，GBDT 节点一般位于数据清洗和特征选择类节点的下游，模型（知识）聚合节点的上游。GBDT 训练得到模型也可以独立存储应用。

5.15.7 HistGBDT

直方提升回归树。这是一种通过直方图技术，优化梯度增强训练算法效率的串行集成算法。使用直方图的方法，将输入数据进行分箱处理（将输入离散化或将输入合并到固定数量的 bin 中），以此进行梯度提升树的生成和训练。在模型性能几乎没有下降的同时，这个方法可以有效的处理缺失值，加快决策树的构建，是一个高效的 Boosting 算法。

5.15.7.1 输入量

HistGBDT 节点输入量如表 143 所示。

表143 HistGBDT 节点输入量列表

名称	类型	描述
in_data	DTTable	带有输入输出变量标识的数据

5.15.7.2 输出量

HistGBDT 节点输出量如表 144 所示。

表144 HistGBDT 节点输出量列表

名称	类型	描述
predictions	DTTable	输出变量真实值、模型预测值及预测误差数据
model	Model	训练后得到的模型
meta_model	MetaModel	模型的超参数



5.15.7.3 配置参数



图271 HistGBDT 节点属性

HistGBDT 节点配置参数如图 271 所示，各参数描述如下：

- 节点名称：节点在建模流程图中的名称标注；
- 节点描述：节点描述信息，用于备注节点所实现的功能；
- 基础配置：指定模型训练算法中的基础参数配置；
- 超参优化：开启超参优化并进入超参优化 Tab 页；
- 交叉验证：指定是否在训练中开启交叉验证；开启后，将在模型详情页显示定性精度指标；

- 高级配置：指定模型训练算法中的高级参数配置。

基础配置中的算法参数介绍如下：

- loss：在提升过程中使用的损失函数；
- learning_rate：基学习器的学习率，与 n_estimators 的大小相互制衡；
- max_iter：基学习器的最大数目，亦是 Boosting 过程的最大迭代次数；
- l2_regularization：L2 正则化参数；参数越多，设置值应该更大[0, 1]。



5.15.7.4 使用说明

在数据建模流程中，HistGBDT 节点一般位于数据清洗和特征选择类节点的下游，模型（知识）聚合节点的上游。HistGBDT 训练得到模型也可以独立存储应用。

5.15.8 KernelRidge

使用核函数的岭回归方法是一种使用核函数将数据集映射到高维空间，再使用岭回归对高维空间的数据进行拟合和预测的方法。在这种方法的预测过程中，只考虑了数据的均值而没有考虑数据的方差。从另一方面来讲，高斯过程回归得到的均值与使用核函数的岭回归相同。适用于中等规模的数据。

5.15.8.1 输入量

KernelRidge 节点输入量如表 145 所示。

表145 KernelRidge 节点输入量列表

名称	类型	描述
in_data	DTTable	带有输入输出变量标识的数据

5.15.8.2 输出量

KernelRidge 节点输出量如表 146 所示。

表146 KernelRidge 节点输出量列表

名称	类型	描述
predictions	DTTable	输出变量真实值、模型预测值及预测误差数据
model	Model	训练后得到的模型
meta_model	MetaModel	模型的超参数



5.15.8.3 配置参数

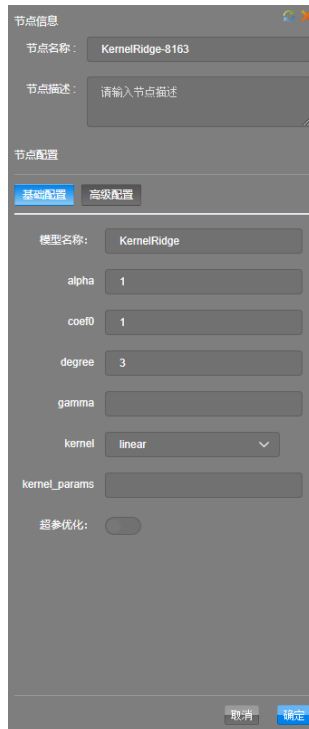


图272 KernelRidge 节点属性

KernelRidge 节点配置参数如图 272 所示，各参数描述如下：

- 节点名称：节点在建模流程图中的名称标注；
- 节点描述：节点描述信息，用于备注节点所实现的功能；
- 基础配置：指定模型训练算法中的基础参数配置；
- 超参优化：开启超参优化并进入超参优化 Tab 页；
- 交叉验证：指定是否在训练中开启交叉验证；开启后，将在模型详情页显示定性精度指标；
- 高级配置：指定模型训练算法中的高级参数配置。

基础配置中的算法参数介绍如下：

- **alpha**：拟合期间添加到核矩阵对角线上的值，并且可以确保计算值形成正定矩阵，防止拟合过程中出现潜在的数值问题；
- **kernel**：核函数。

5.15.8.4 使用说明

在数据建模流程中，KernelRidge 节点一般位于数据清洗和特征选择类节点



的下游，模型（知识）聚合节点的上游。KernelRidge 训练得到模型也可以独立存储应用。

5.15.9 KNN

K 近邻回归算法是一种通过距离的计算找到近邻点，通过近邻点的信息进行预测的方法。

5.15.9.1 输入量

KNN 节点输入量如表 147 所示。

表147 KNN 节点输入量列表

名称	类型	描述
in_data	DTTable	带有输入输出变量标识的数据

5.15.9.2 输出量

KNN 节点输出量如表 148 所示。

表148 KNN 节点输出量列表

名称	类型	描述
predictions	DTTable	输出变量真实值、模型预测值及预测误差数据
model	Model	训练后得到的模型
meta_model	MetaModel	模型的超参数



5.15.9.3 配置参数

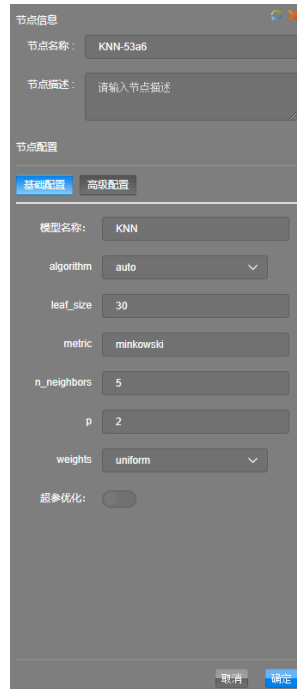


图273 KNN 节点属性

KNN 节点配置参数如图 273 所示，各参数描述如下：

- 节点名称：节点在建模流程图中的名称标注；
- 节点描述：节点描述信息，用于备注节点所实现的功能；
- 基础配置：指定模型训练算法中的基础参数配置；
- 超参优化：开启超参优化并进入超参优化 Tab 页；
- 交叉验证：指定是否在训练中开启交叉验证；开启后，将在模型详情页

显示定性精度指标；

- 高级配置：指定模型训练算法中的高级参数配置。

基础配置中的算法参数介绍如下：

- `n_neighbors`：预测时使用的近邻点个数；
- `weights`：预测时的权重函数；
- `p`：闵可夫斯基距离（Minkowski distance）的阶数。

5.15.9.4 使用说明

在数据建模流程中，KNN 节点一般位于数据清洗和特征选择类节点的下游，模型（知识）聚合节点的上游。KNN 训练得到模型也可以独立存储应用。



5.15.10 MLP

多层感知器（MLP，Multilayer Perceptron）是一种前馈人工神经网络模型，其将输入的多个数据集映射到单一的输出的数据集上。这是一种神经网络的方法，使用隐藏层进行线性关系的计算，使用激活函数将线性关系非线性化；并通过反向传播的方式更新权重，从而实现模型的优化。

5.15.10.1 输入量

MLP 节点输入量如表 149 所示。

表149 MLP 节点输入量列表

名称	类型	描述
in_data	DTTable	带有输入输出变量标识的数据

5.15.10.2 输出量

MLP 节点输出量如表 150 所示。

表150 MLP 节点输出量列表

名称	类型	描述
predictions	DTTable	输出变量真实值、模型预测值及预测误差数据
model	Model	训练后得到的模型
meta_model	MetaModel	模型的超参数



5.15.10.3 配置参数

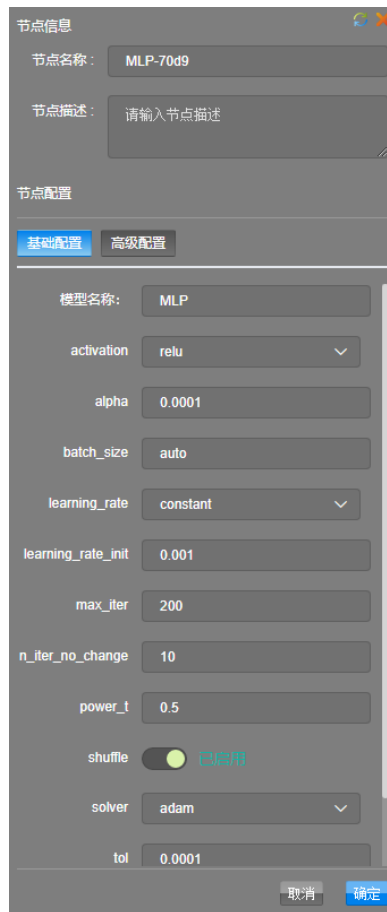


图274 MLP 节点属性

MLP 节点配置参数如图 274 所示，各参数描述如下：

- 节点名称：节点在建模流程图中的名称标注；
- 节点描述：节点描述信息，用于备注节点所实现的功能；
- 基础配置：指定模型训练算法中的基础参数配置；
- 超参优化：开启超参优化并进入超参优化 Tab 页；
- 交叉验证：指定是否在训练中开启交叉验证；开启后，将在模型详情页

显示定性精度指标；

- 高级配置：指定模型训练算法中的高级参数配置。

基础配置中的算法参数介绍如下：

- hidden_layer_sizes：定义了隐藏层的层数及神经元个数；
- activation：激活函数；



- **solver**: 权重优化函数 ('lbfgs': 基于拟牛顿法, 'sgd': 随机梯度下降, 'adam': 基于随机梯度的优化器);
- **max_iter**: 最大迭代次数。

5.15.10.4 使用说明

在数据建模流程中，MLP 节点一般位于数据清洗和特征选择类节点的下游，模型（知识）聚合节点的上游。MLP 训练得到模型也可以独立存储应用。

5.15.11 RNN

限定半径最近邻回归树，是一种通过距离的计算找到近邻点，通过近邻点的信息进行预测的方法。与 KNN 不同的是，RNN 通过设定各点到其的距离来找到近邻点，其个数不是确定的；而 KNN 的近邻点个数是确定的。

5.15.11.1 输入量

RNN 节点输入量如表 151 所示。

表151 RNN 节点输入量列表

名称	类型	描述
in_data	DTTable	带有输入输出变量标识的数据

5.15.11.2 输出量

RNN 节点输出量如表 152 所示。

表152 RNN 节点输出量列表

名称	类型	描述
predictions	DTTable	输出变量真实值、模型预测值及预测误差数据
model	Model	训练后得到的模型
meta_model	MetaModel	模型的超参数



5.15.11.3 配置参数



图275 RNN 节点属性

RNN 节点配置参数如图 275 所示，各参数描述如下：

- 节点名称：节点在建模流程图中的名称标注；
- 节点描述：节点描述信息，用于备注节点所实现的功能；
- 基础配置：指定模型训练算法中的基础参数配置；
- 超参优化：开启超参优化并进入超参优化 Tab 页；
- 交叉验证：指定是否在训练中开启交叉验证；开启后，将在模型详情页

显示定性精度指标；

- 高级配置：指定模型训练算法中的高级参数配置。

基础配置中的算法参数介绍如下：

- radius：定义查询的参数空间范围；
- weights：预测时的权重函数；
- p：闵可夫斯基距离（Minkowski distance）的阶数。



5.15.11.4 使用说明

在数据建模流程中，RNN 节点一般位于数据清洗和特征选择类节点的下游，模型（知识）聚合节点的上游。RNN 训练得到模型也可以独立存储应用。

5.15.12 RandomForest

随机森林，是一种对样本进行随机抽样，并构造多个弱学习器的并行集成学习方法。通过对所有样本进行有放回的随机抽样，对每个样本集分别训练一个基学习器，最后将多个基学习器的结果按相同权重进行组合，得到最终的拟合或预测结果。

5.15.12.1 输入量

RandomForest 节点输入量如表 153 所示。

表153 RandomForest 节点输入量列表

名称	类型	描述
in_data	DTTable	带有输入输出变量标识的数据

5.15.12.2 输出量

RandomForest 节点输出量如表 154 所示。

表154 RandomForest 节点输出量列表

名称	类型	描述
predictions	DTTable	输出变量真实值、模型预测值及预测误差数据
model	Model	训练后得到的模型
meta_model	MetaModel	模型的超参数



5.15.12.3 配置参数

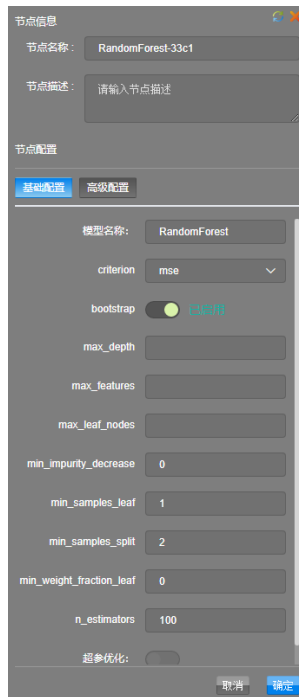


图276 RandomForest 节点属性

RandomForest 节点配置参数如图 276 所示，各参数描述如下：

- 节点名称：节点在建模流程图中的名称标注；
- 节点描述：节点描述信息，用于备注节点所实现的功能；
- 基础配置：指定模型训练算法中的基础参数配置；
- 超参优化：开启超参优化并进入超参优化 Tab 页；
- 交叉验证：指定是否在训练中开启交叉验证；开启后，将在模型详情页显示定性精度指标；

- 高级配置：指定模型训练算法中的高级参数配置。

基础配置中的算法参数介绍如下：

- `n_estimators`：基学习器的最大数目；
- `criterion`：测量节点分割质量的函数；
- `bootstrap`：是否对所有样本进行有放回的抽样，若为 `False`，使用全部样本进行训练；
- `oob_score`：误差计算方式，是否使用袋外的样本计算 `R2` 指标。



5.15.12.4 使用说明

在数据建模流程中，RandomForest 节点一般位于数据清洗和特征选择类节点的下游，模型（知识）聚合节点的上游。RandomForest 训练得到模型也可以独立存储应用。

5.15.13 SVR

支持向量机回归。这是一种通过核函数将数据映射到高维线性空间，再使用支持向量的思想选择最佳的超平面（高维空间上的线性函数），再通过超平面进行预测的方法。支持向量机回归与线性回归相比，支持向量回归表示只要在虚线内部的值都可认为是预测正确，只要计算虚线外部的值的损失即可。和使用核函数的岭回归相比，正则化过程被支持向量的思想替代，训练速度也比 KernelRidge 更快，在数据集大小没有要求，但其精度可能比 KernelRidge 要低。

5.15.13.1 输入量

SVR 节点输入量如表 155 所示。

表155 SVR 节点输入量列表

名称	类型	描述
in_data	DTTable	带有输入输出变量标识的数据

5.15.13.2 输出量

SVR 节点输出量如表 156 所示。

表156 SVR 节点输出量列表

名称	类型	描述
predictions	DTTable	输出变量真实值、模型预测值及预测误差数据
model	Model	训练后得到的模型
meta_model	MetaModel	模型的超参数



5.15.13.3 配置参数



图277 SVR 节点属性

SVR 节点配置参数如图 277 所示，各参数描述如下：

- 节点名称：节点在建模流程图中的名称标注；
- 节点描述：节点描述信息，用于备注节点所实现的功能；
- 基础配置：指定模型训练算法中的基础参数配置；
- 超参优化：开启超参优化并进入超参优化 Tab 页；
- 交叉验证：指定是否在训练中开启交叉验证；开启后，将在模型详情页

显示定性精度指标；

- 高级配置：指定模型训练算法中的高级参数配置。

基础配置中的算法参数介绍如下：

- kernel：核函数；
- C：误差项的惩罚参数。

5.15.13.4 使用说明

在数据建模流程中，SVR 节点一般位于数据清洗和特征选择类节点的下游，模型（知识）聚合节点的上游。SVR 训练得到模型也可以独立存储应用。



5.15.14 EnsembleRBF

5.15.14.1 输入量

EnsembleRBF 节点输入量如表 157 所示。

表157 EnsembleRBF 节点输入量列表

名称	类型	描述
in_data	DTTable	带有输入输出变量标识的数据

5.15.14.2 输出量

EnsembleRBF 节点输出量如表 158 所示。

表158 EnsembleRBF 节点输出量列表

名称	类型	描述
predictions	DTTable	输出变量真实值、模型预测值及预测误差数据
model	Model	训练后得到的模型
meta_model	MetaModel	模型的超参数

5.15.14.3 配置参数

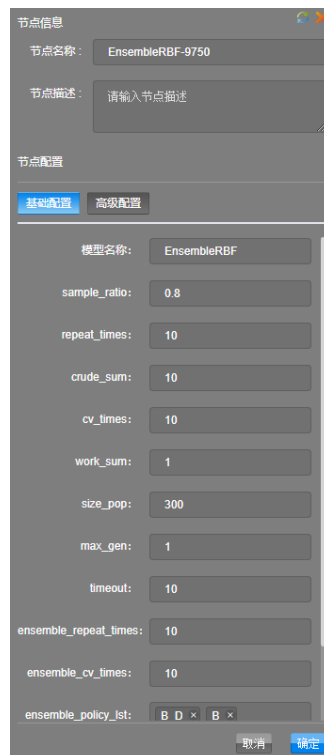


图278 EnsembleRBF 节点属性



EnsembleRBF 节点配置参数如图 278 所示，各参数描述如下：

- 节点名称：节点在建模流程图中的名称标注；
- 节点描述：节点描述信息，用于备注节点所实现的功能；
- 基础配置：指定模型训练算法中的基础参数配置；
- 超参优化：开启超参优化并进入超参优化 Tab 页；
- 交叉验证：指定是否在训练中开启交叉验证；开启后，将在模型详情页显示定性精度指标；
- 高级配置：指定模型训练算法中的高级参数配置。

5.15.14.4 使用说明

在数据建模流程中，EnsembleRBF 节点一般位于数据清洗和特征选择类节点的下游，模型（知识）聚合节点的上游。EnsembleRBF 训练得到模型也可以独立存储应用。

5.15.15 RBFsSurrogate

5.15.15.1 输入量

RBFsSurrogate 节点输入量如表 159 所示。

表159 RBFsSurrogate 节点输入量列表

名称	类型	描述
in_data	DTTable	带有输入输出变量标识的数据

5.15.15.2 输出量

RBFsSurrogate 节点输出量如表 160 所示。

表160 RBFsSurrogate 节点输出量列表

名称	类型	描述
predictions	DTTable	输出变量真实值、模型预测值及预测误差数据
model	Model	训练后得到的模型
meta_model	MetaModel	模型的超参数



5.15.15.3 配置参数



图279 RBFSurrogate 节点属性

RBFSurrogate 节点配置参数如图 279 所示，各参数描述如下：

- 节点名称：节点在建模流程图中的名称标注；
- 节点描述：节点描述信息，用于备注节点所实现的功能；
- 基础配置：指定模型训练算法中的基础参数配置；
- 超参优化：开启超参优化并进入超参优化 Tab 页；
- 交叉验证：指定是否在训练中开启交叉验证；开启后，将在模型详情页显示定性精度指标；
- 高级配置：指定模型训练算法中的高级参数配置。

5.15.15.4 使用说明

在数据建模流程中，RBFSurrogate 节点一般位于数据清洗和特征选择类节点的下游，模型（知识）聚合节点的上游。RBFSurrogate 训练得到模型也可以独立存储应用。

5.15.16 多项式拟合

基于不同阶数的多项式进行模型训练。



5.15.16.1 输入量

多项式拟合节点输入量如表 161 所示。

表161 多项式拟合节点输入量列表

名称	类型	描述
in_data	DTTable	带有输入输出变量标识的数据

5.15.16.2 输出量

多项式拟合节点输出量如表 162 所示。

表162 多项式拟合节点输出量列表

名称	类型	描述
predictions	DTTable	输出变量真实值、模型预测值及预测误差数据
model	Model	训练后得到的模型
meta_model	MetaModel	模型的超参数

5.15.16.3 配置参数



图280 多项式拟合节点属性

多项式拟合节点配置参数如图 280 所示，各参数描述如下：

- 节点名称：节点在建模流程图中的名称标注；
- 节点描述：节点描述信息，用于备注节点所实现的功能；
- 模型名称：训练得到模型的名称；
- 阶数：多项式拟合时的最高阶数。



5.15.16.4 使用说明

在数据建模流程中，多项式拟合节点一般位于数据清洗和特征选择类节点的下流，模型（知识）聚合节点的上游。多项式拟合训练得到模型也可以独立存储应用。

5.16 调度策略

5.16.1 EnsembleBagging

调度策略。

5.16.1.1 输入量

EnsembleBagging 节点输入量如表 163 所示。

表163 EnsembleBagging 节点输入量列表

名称	类型	描述
in_data	DTTable	带有输入输出变量标识的数据

5.16.1.2 输出量

EnsembleBagging 节点输出量如表 164 所示。

表164 EnsembleBagging 节点输出量列表

名称	类型	描述
predictions	DTTable	输出变量真实值、模型预测值及预测误差数据
model	Model	训练后得到的模型
meta_model	MetaModel	模型的超参数



5.16.1.3 配置参数



图281 EnsembleBagging 节点属性

EnsembleBagging 节点配置参数如图 281 所示，各参数描述如下：

- 节点名称：节点在建模流程图中的名称标注；
- 节点描述：节点描述信息，用于备注节点所实现的功能；
- 基础配置：指定模型训练算法中的基础参数配置；
- 超参优化：开启超参优化并进入高级配置；
- 高级配置：指定模型训练算法中的高级参数配置。

5.16.1.4 使用说明

在数据建模流程中，EnsembleBagging 节点一般位于数据清洗和特征选择类节点的下游，模型（知识）聚合节点的上游。EnsembleBagging 训练得到模型也可以独立存储应用。



5.16.2 HierarchicalStratify

调度策略。

5.16.2.1 输入量

HierarchicalStratify 节点输入量如表 165 所示。

表165 HierarchicalStratify 节点输入量列表

名称	类型	描述
in_data	DTTable	带有输入输出变量标识的数据

5.16.2.2 输出量

HierarchicalStratify 节点输出量如表 166 所示。

表166 HierarchicalStratify 节点输出量列表

名称	类型	描述
predictions	DTTable	输出变量真实值、模型预测值及预测误差数据
model	Model	训练后得到的模型
meta_model	MetaModel	模型的超参数

5.16.2.3 配置参数



图282 HierarchicalStratify 节点属性

HierarchicalStratify 节点配置参数如图 282 所示，各参数描述如下：



- 节点名称：节点在建模流程图中的名称标注；
- 节点描述：节点描述信息，用于备注节点所实现的功能；
- 基础配置：指定模型训练算法中的基础参数配置；
- 超参优化：开启超参优化并进入高级配置；
- 高级配置：指定模型训练算法中的高级参数配置。

5.16.2.4 使用说明

在数据建模流程中，HierachicalStratify 节点一般位于数据清洗和特征选择类节点的下游，模型（知识）聚合节点的上游。HierachicalStratify 训练得到模型也可以独立存储应用。

5.16.3 OTSpecifiedSDASRegressor

调度策略。

5.16.3.1 输入量

OTSpecifiedSDASRegressor 节点输入量如表 167 所示。

表167 OTSpecifiedSDASRegressor 节点输入量列表

名称	类型	描述
in_data	DTTable	带有输入输出变量标识的数据

5.16.3.2 输出量

OTSpecifiedSDASRegressor 节点输出量如表 168 所示。

表168 OTSpecifiedSDASRegressor 节点输出量列表

名称	类型	描述
predictions	DTTable	输出变量真实值、模型预测值及预测误差数据
model	Model	训练后得到的模型
meta_model	MetaModel	模型的超参数



5.16.3.3 配置参数



图283 OTSpecifiedSDASRegressor 节点属性

OTSpecifiedSDASRegressor 节点配置参数如图 283 所示，各参数描述如下：

- 节点名称：节点在建模流程图中的名称标注；
- 节点描述：节点描述信息，用于备注节点所实现的功能；
- 基础配置：指定模型训练算法中的基础参数配置；
- 超参优化：开启超参优化并进入高级配置；
- 高级配置：指定模型训练算法中的高级参数配置。

5.16.3.4 使用说明

在数据建模流程中，OTSpecifiedSDASRegressor 节点一般位于数据清洗和特征选择类节点的下游，模型（知识）聚合节点的上游。OTSpecifiedSDASRegressor 训练得到模型也可以独立存储应用。



5.17 机理模型

5.17.1 显函数

显函数节点支持用户指定包含未知参数的映射函数，通常情况下这些映射函数往往由工程经验或专家知识提供。显函数节点能够基于训练数据，对未知参数进行优化，以得到最匹配当前应用场景的模型。

5.17.1.1 输入量

显函数节点输入量如表 169 所示。

表169 显函数节点输入量列表

名称	类型	描述
in_data	DTTable	带有输入输出变量标识的数据

5.17.1.2 输出量

显函数节点输出量如表 170 所示。

表170 显函数节点输出量列表

名称	类型	描述
predictions	DTTable	输出变量真实值、模型预测值及预测误差数据
model	Model	训练后得到的机理模型
meta_model	MetaModel	机理模型的超参数



5.17.1.3 配置参数



图284 显函数节点属性

显函数节点配置参数如图 284 所示，各参数描述如下：

- 节点名称：节点在建模流程图中的名称标注；
- 节点描述：节点描述信息，用于备注节点所实现的功能；
- 拟合参数：待求解的参数；
- 表达式：输出变量计算的机理模型公式（机理模型）；
- 优化算法：参数拟合过程所使用的优化算法；
- 结果：未知参数的寻优结果，单击查看按钮，即可查看所有拟合参数的拟合结果，如图 285 所示。

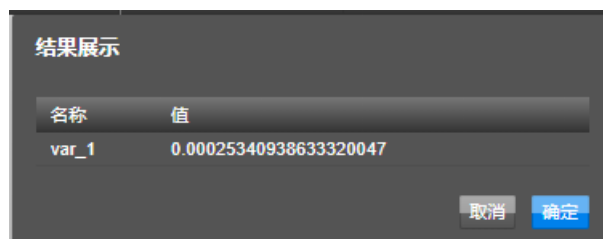


图285 拟合参数结果

需要注意的是，当该节点未执行时，需要拟合的参数和机理模型公式对于该节点是未知的。当该节点被执行时，系统会自动弹出机理模型配置窗口，用户在弹窗中对待拟合的机理模型进行配置，如图 286 所示。





图286 机理模型

如图 286 所示，机理模型弹窗左侧为待拟合的参数，用户通过新建按钮添加带单位和参数，并在列表中对参数范围和初值进行设定。用户通过右侧的变量名称、数学运算符和基础数学函数实现计算公式的配置。

需要注意的是，输出变量数量必须与传变量 `in_data` 输出变量数量相同，每个输出变量的计算公式分行排列。此外，输出变量不支持出现在等式的右侧。

5.17.1.4 使用说明

在数据建模流程中，显函数节点一般位于数据清洗和特征选择类节点的下游，模型（知识）聚合节点的上游。经显函数训练得到模型也可以独立存储应用。

5.18 时序训练

5.18.1 ARIMA

ARIMA 是一种自回归差分移动平均的线性时序预测方法，其通过拟合时间序列的历史数据的潜在规律进行未来的预测。

5.18.1.1 输入量

ARIMA 节点输入量如表 171 所示。

表171 ARIMA 节点输入量列表

名称	类型	描述
<code>in_data</code>	TS-DTTable	带有时间变量与预测变量标识的时序数据



5.18.1.2 输出量

ARIMA 节点输出量如表 172 所示。

表172 ARIMA 节点输出量列表

名称	类型	描述
tsmodel	TS-Model	训练后得到的时序模型

5.18.1.3 配置参数



图287 ARIMA 节点属性

ARIMA 节点配置参数如图 287 所示，各参数描述如下：

- 节点名称：节点在建模流程图中的名称标注；
- 节点描述：节点描述信息，用于备注节点所实现的功能；
- 基础配置：指定时序模型训练算法中的基础参数配置；
- 超参优化：开启超参优化并进入超参优化 Tab 页。

基础配置中的算法参数介绍如下：

- p：自回归阶数；
- d：差分阶数；
- q：移动平均阶数。



需要注意的是，当该节点未执行时，数据信息对于该节点是未知的。当该节点被执行时，系统会自动弹出可视化配置面板，用户可在弹窗中实现算法参数的设定操作，如图 288 所示。

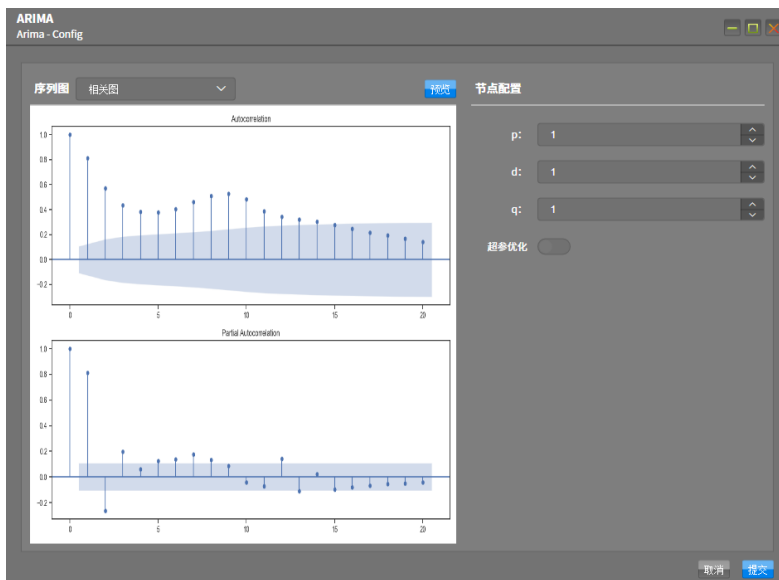


图288 ARIMA 节点可视化配置面板

在 ARIMA 节点可视化配置面板中，右侧为节点参数配置，左侧为根据设定的参数计算得到的可视化图形，包括序列图、差分序列图以及相关图等。通过顶部的下拉菜单可以切换可视化图形，以及通过预览功能更新计算结果。

点击开始流程并成功运行后，于左侧资源树下可查看该节点的输出模型，如图 289 所示。

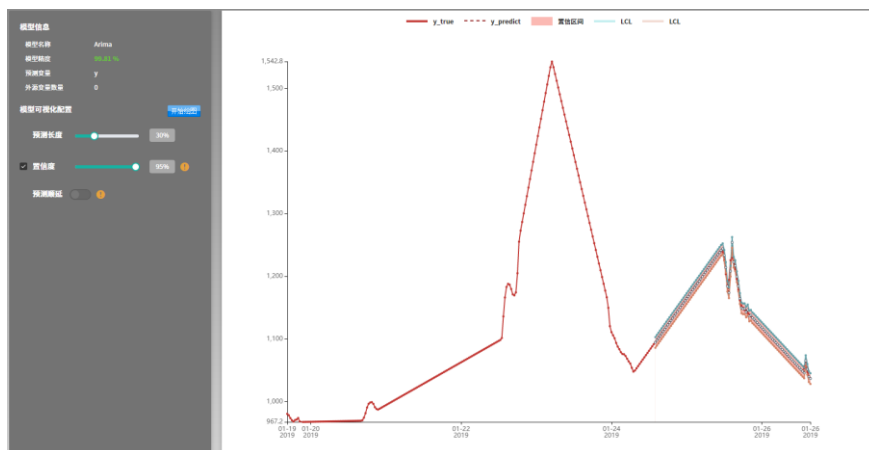


图289 ARIMA 模型详情



在模型详情页面，可查看对应的算法名称、模型精度、预测变量名称、外源变量个数等信息。同时于画面右侧显示在训练集上的预测结果，并支持交互式的可视化配置，包括如下可配置参数：

- 预测长度：从训练样本末尾向前截取对应长度（以占整体样本数量的百分数表示）的数据作为测试集；
- 置信度：开启后，将根据设置的置信度，在输出预测值的基础上同时输出置信区间，包括置信下限（LCL）与置信上限（UCL）；
- 预测顺延：开启后，从第一个预测样本点开始停止更新观测值，以作为严格的预测对象。

5.18.1.4 使用说明

在数据建模流程中，ARIMA 节点一般位于时序变量设定或时序预处理节点的下流，时序预测节点的上游。ARIMA 训练得到模型也可以独立存储应用。

ARIMA 节点支持在时序变量设定中引入外源变量。

5.18.2 SARIMA

SARIMA 是一种季节性自回归差分移动平均的线性时序预测方法，其通过拟合时间序列的历史数据的潜在规律进行未来的预测，并在 ARIMA 的基础上考虑了周期性因素。

5.18.2.1 输入量

SARIMA 节点输入量如表 173 所示。

表173 SARIMA 节点输入量列表

名称	类型	描述
in_data	TS-DTTable	带有时间变量与预测变量标识的时序数据

5.18.2.2 输出量

SARIMA 节点输出量如表 174 所示。

表174 SARIMA 节点输出量列表

名称	类型	描述
tsmodel	TS-Model	训练后得到的时序模型



5.18.2.3 配置参数



图290 SARIMA 节点属性

SARIMA 节点配置参数如图 290 所示，各参数描述如下：

- 节点名称：节点在建模流程图中的名称标注；
- 节点描述：节点描述信息，用于备注节点所实现的功能；
- 基础配置：指定时序模型训练算法中的基础参数配置；
- 超参优化：开启超参优化并进入超参优化 Tab 页。

基础配置中的算法参数介绍如下：

- p：自回归阶数；
- d：差分阶数；
- q：移动平均阶数；
- P：季节性自回归阶数；
- D：季节性差分阶数；
- Q：季节性移动平均阶数；
- S：时间序列周期。



需要注意的是，当该节点未执行时，数据信息对于该节点是未知的。当该节点被执行时，系统会自动弹出可视化配置面板，用户可在弹窗中实现算法参数的设定操作，如图 291 所示。

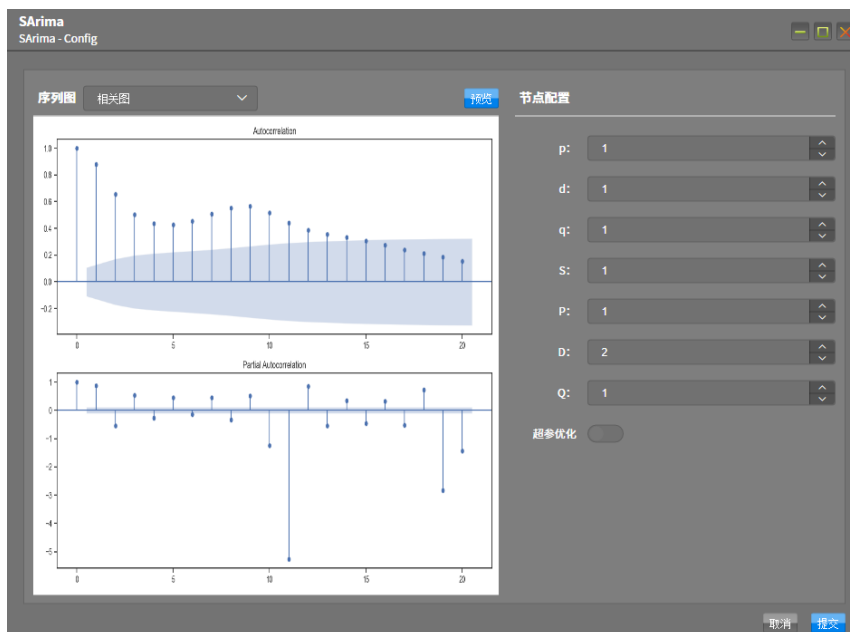


图291 SARIMA 节点可视化配置面板

在 SARIMA 节点可视化配置面板中，右侧为节点参数配置，左侧为根据设定的参数计算得到的可视化图形，包括序列图、差分序列图以及相关图等。通过顶部的下拉菜单可以切换可视化图形，以及通过预览功能更新计算结果。

点击开始流程并成功运行后，于左侧资源树下可查看该节点的输出模型，如图 292 所示。

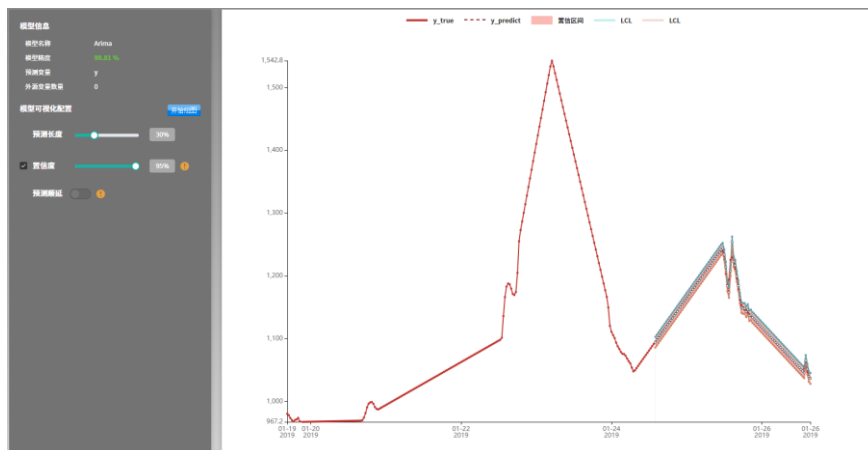


图292 SARIMA 模型详情



在模型详情页面，可查看对应的算法名称、模型精度、预测变量名称、外源变量个数等信息。同时于画面右侧显示在训练集上的预测结果，并支持交互式的可视化配置，包括如下可配置参数：

- 预测长度：从训练样本末尾向前截取对应长度（以占整体样本数量的百分数表示）的数据作为测试集；
- 置信度：开启后，将根据设置的置信度，在输出预测值的基础上同时输出置信区间，包括置信下限（LCL）与置信上限（UCL）；
- 预测顺延：开启后，从第一个预测样本点开始停止更新观测值，以作为严格的预测对象。

5.18.2.4 使用说明

在数据建模流程中，SARIMA 节点一般位于时序变量设定或时序预处理节点的下流，时序预测节点的上游。SARIMA 训练得到模型也可以独立存储应用。

SARIMA 节点支持在时序变量设定中引入外源变量。

5.18.3 ES

ES 实现了 1-3 次指数平滑算法，随着阶数的增加，指数平滑算法逐渐开始考虑时间序列中的趋势项（2 次）和季节项（3 次）。需要注意的是，指数平滑本质是一种时序平滑手段，适用的场景比较有限，做长窗预测的能力存在较大局限。

5.18.3.1 输入量

ES 节点输入量如表 175 所示。

表175 ES 节点输入量列表

名称	类型	描述
in_data	TS-DTTable	带有时间变量与预测变量标识的时序数据

5.18.3.2 输出量

ES 节点输出量如表 176 所示。

表176 ES 节点输出量列表

名称	类型	描述
tmodel	TS-Model	训练后得到的时序模型



5.18.3.3 配置参数



图293 ES 节点属性

ES 节点配置参数如图 293 所示，各参数描述如下：

- 节点名称：节点在建模流程图中的名称标注；
- 节点描述：节点描述信息，用于备注节点所实现的功能；
- 基础配置：指定时序模型训练算法中的基础参数配置；
- 高级配置：指定模型训练算法中的高级参数配置。

基础配置中的算法参数介绍如下：

- trend：趋势分量的计算方法；
- seasonal：季节性分量的计算方法。

点击开始流程并成功运行后，于左侧资源树下可查看该节点的输出模型，如图 294 所示。



图294 ES 模型详情



在模型详情页面，可查看对应的算法名称、模型精度、预测变量名称、外源变量个数等信息。同时于画面右侧显示在训练集上的预测结果，并支持交互式的可视化配置，包括如下可配置参数：

- 预测长度：从训练样本末尾向前截取对应长度（以占整体样本数量的百分数表示）的数据作为测试集；
- 置信度：ES 算法不支持计算置信区间；
- 预测顺延：开启后，从第一个预测样本点开始停止更新观测值，以作为严格的预测对象。

5.18.3.4 使用说明

在数据建模流程中，ES 节点一般位于时序变量设定或时序预处理节点的下流，时序预测节点的上游。ES 训练得到模型也可以独立存储应用。

ES 节点不支持在时序变量设定中引入外源变量。

5.18.4 LOESS

LOESS 是一种通过构造多个时间窗口并将其加权平均进行预测的方法，这种方法主要用于拟合序列中的变化趋势、平滑、异常点筛选。需要注意的是，LOESS 本质是一种时序平滑手段，适用的场景比较有限，做长窗预测的能力存在较大局限。

5.18.4.1 输入量

LOESS 节点输入量如表 177 所示。

表177 LOESS 节点输入量列表

名称	类型	描述
in_data	TS-DTTable	带有时间变量与预测变量标识的时序数据

5.18.4.2 输出量

LOESS 节点输出量如表 178 所示。

表178 LOESS 节点输出量列表

名称	类型	描述
tsmodel	TS-Model	训练后得到的时序模型



5.18.4.3 配置参数



图295 LOESS 节点属性

LOESS 节点配置参数如图 295 所示，各参数描述如下：

- 节点名称：节点在建模流程图中的名称标注；
- 节点描述：节点描述信息，用于备注节点所实现的功能；
- 基础配置：指定时序模型训练算法中的基础参数配置；
- 高级配置：指定模型训练算法中的高级参数配置。

基础配置中的算法参数介绍如下：

- local_window_size：局部时间窗口宽度；
- degree：模型的阶数，如 1 表示模型是一阶线性的。

点击开始流程并成功运行后，于左侧资源树下可查看该节点的输出模型，如图 296 所示。

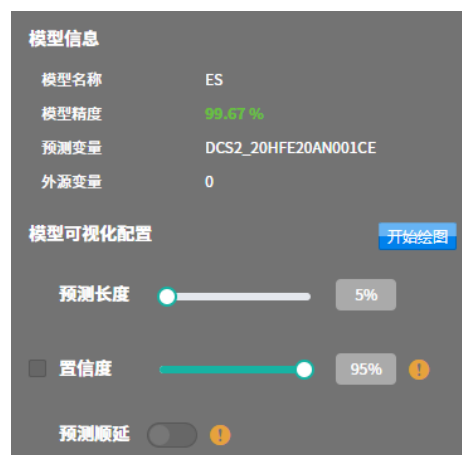


图296 LOESS 模型详情



在模型详情页面，可查看对应的算法名称、模型精度、预测变量名称、外源变量个数等信息。同时于画面右侧显示在训练集上的预测结果，并支持交互式的可视化配置，包括如下可配置参数：

- 预测长度：从训练样本末尾向前截取对应长度（以占整体样本数量的百分数表示）的数据作为测试集；
- 置信度：LOESS 算法不支持计算置信区间；
- 预测顺延：开启后，从第一个预测样本点开始停止更新观测值，以作为严格的预测对象。

5.18.4.4 使用说明

在数据建模流程中，LOESS 节点一般位于时序变量设定或时序预处理节点的下流，时序预测节点的上游。LOESS 训练得到模型也可以独立存储应用。

LOESS 节点不支持在时序变量设定中引入外源变量。

5.18.5 Prophet

Prophet 是一种通过专业知识进行时序预测的方法，其中专业知识包括实际生活中的假期/周期/变化点信息。这种方法采用了时间序列分解的思想，通过将时间序列分解为趋势/季节/假期/误差项进行时序预测，分析人员可以通过设定详细的专业信息，以达到对时间序列的精确预测。这个方法对周期尺度较大（至少为/日）的数据具有较高的预测精度，但对于周期尺度和尖锐点尺度较小的时间序列（/秒），它的预测效果可能不够好。

5.18.5.1 输入量

Prophet 节点输入量如表 179 所示。

表179 Prophet 节点输入量列表

名称	类型	描述
in_data	TS-DTTable	带有时间变量与预测变量标识的时序数据

5.18.5.2 输出量

Prophet 节点输出量如表 180 所示。

表180 Prophet 节点输出量列表



名称	类型	描述
tsmodel	TS-Model	训练后得到的时序模型

5.18.5.3 配置参数



图297 Prophet 节点属性

Prophet 节点配置参数如图 297 所示，各参数描述如下：

- 节点名称：节点在建模流程图中的名称标注；
- 节点描述：节点描述信息，用于备注节点所实现的功能；
- 基础配置：指定时序模型训练算法中的基础参数配置；
- 高级配置：指定模型训练算法中的高级参数配置。

基础配置中的算法参数介绍如下：

- growth：趋势项类型；
- seasonality_mode：季节项分解和组合的方式；
- n_changepoints：设定潜在变化点的数目。

点击开始流程并成功运行后，于左侧资源树下可查看该节点的输出模型，如图 298 所示。





图298 Prophet 模型详情

在模型详情页面，可查看对应的算法名称、模型精度、预测变量名称、外源变量个数等信息。同时于画面右侧显示在训练集上的预测结果，并支持交互式的可视化配置，包括如下可配置参数：

- 预测长度：从训练样本末尾向前截取对应长度（以占整体样本数量的百分数表示）的数据作为测试集；
- 置信度：开启后，将根据设置的置信度，在输出预测值的基础上同时输出置信区间，包括置信下限（LCL）与置信上限（UCL）；
- 预测顺延：开启后，从第一个预测样本点开始停止更新观测值，以作为严格的预测对象。

5.18.5.4 使用说明

在数据建模流程中，Prophet 节点一般位于时序变量设定或时序预处理节点的下游，时序预测节点的上游。Prophet 训练得到模型也可以独立存储应用。

Prophet 节点不支持在时序变量设定中引入外源变量。

5.18.6 TBATS

TBATS 是一种自动化模型，其在 ARIMA 的基础上更进一步，能够同时识别不同的季节性模式。但和 ARIMA 相比有个弊端：ARIMA 能引入外源变量，而 TBATS 只能预测一个变量。

5.18.6.1 输入量

TBATS 节点输入量如表 181 所示。



表181 TBATS 节点输入量列表

名称	类型	描述
in_data	TS-DTTable	带有时间变量与预测变量标识的时序数据

5.18.6.2 输出量

TBATS 节点输出量如表 182 所示。

表182 TBATS 节点输出量列表

名称	类型	描述
tsmodel	TS-Model	训练后得到的时序模型

5.18.6.3 配置参数

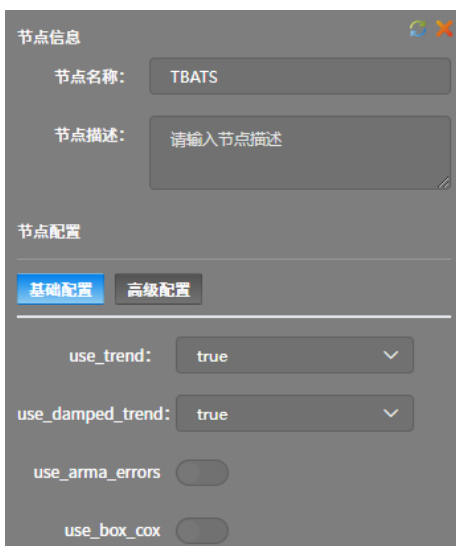


图299 TBATS 节点属性

TBATS 节点配置参数如图 299 所示，各参数描述如下：

- 节点名称：节点在建模流程图中的名称标注；
- 节点描述：节点描述信息，用于备注节点所实现的功能；
- 基础配置：指定时序模型训练算法中的基础参数配置；
- 高级配置：指定模型训练算法中的高级参数配置。

基础配置中的算法参数介绍如下：

- use_trend：是否引入趋势项；
- use_damped_trend：是否使用阻尼趋势；



- `use_arma_errors`: 是否使用 ARMA 模型对残差进行建模，设置为 False 会极大程度上增加计算速度；
- `use_box_cox`: 是否进行 Box-Cox 变换。

点击开始流程并成功运行后，于左侧资源树下可查看该节点的输出模型，如图 298 所示。

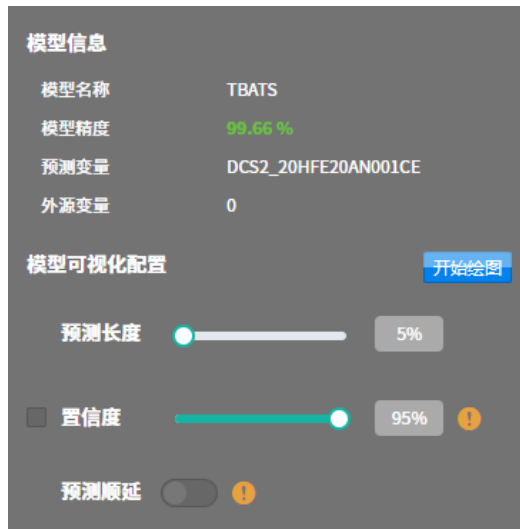


图300 TBATS 模型详情

在模型详情页面，可查看对应的算法名称、模型精度、预测变量名称、外源变量个数等信息。同时于画面右侧显示在训练集上的预测结果，并支持交互式的可视化配置，包括如下可配置参数：

- **预测长度**：从训练样本末尾向前截取对应长度（以占整体样本数量的百分数表示）的数据作为测试集；
- **置信度**：开启后，将根据设置的置信度，在输出预测值的基础上同时输出置信区间，包括置信下限（LCL）与置信上限（UCL）；需要说明的是，由于 TBATS 算法计算较为耗时，因此在模型详情页屏蔽了该功能；
- **预测顺延**：开启后，从第一个预测样本点开始停止更新观测值，以作为严格的预测对象；需要说明的是，由于 TBATS 算法计算较为耗时，因此在模型详情页屏蔽了该功能。



5.18.6.4 使用说明

在数据建模流程中，TBATS 节点一般位于时序变量设定或时序预处理节点的下流，时序预测节点的上游。TBATS 训练得到模型也可以独立存储应用。

TBATS 节点不支持在时序变量设定中引入外源变量。

5.19 模型操作

DTEmpower 提供了对建模流程训练得到的模型进行聚合、性能评估、保存和读取等多种作用于模型的功能节点。

5.19.1 模型聚合

聚合节点支持将机理模型、回归模型、AI Agent 模型等多种模型聚合成新的模型，聚合后的模型输入输出变量均为聚合前各模型输入输出变量的并集。

5.19.1.1 输入量

聚合节点输入量如表 183 所示。

表183 聚合节点输入量列表

名称	类型	描述
models	List	待聚合的多个模型组成的列表

5.19.1.2 输出量

聚合节点输出量如表 184 所示。

表184 聚合节点输出量列表

名称	类型	描述
model	Model	聚合后的模型



5.19.1.3 配置参数

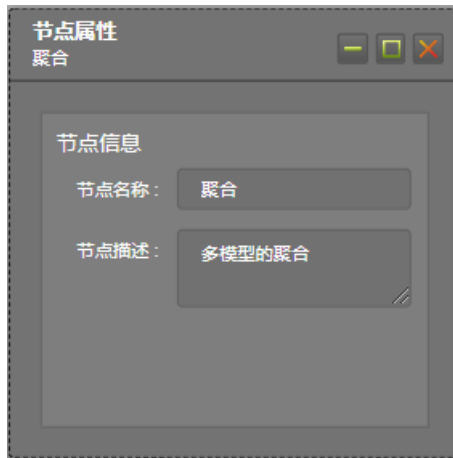


图301 聚合节点属性

聚合节点配置参数如图 301 所示，各参数描述如下：

- 节点名称：节点在建模流程图中的名称标注；
- 节点描述：节点描述信息，用于备注节点所实现的功能。

5.19.1.4 使用说明

在数据建模流程中，聚合节点一般位于各模型训练节点的下游，其输出的模型支持单独存储和使用。

5.19.2 模型对比

基于同一测试数据集，对（多个）模型性能进行评估，输出包含 MSE、R2 等模型评价指标在内的模型测试结果数据。

5.19.2.1 输入量

对比节点输入量如表 185 所示。

表185 对比节点输入量列表

名称	类型	描述
models	List	待聚合的多个模型组成的列表

5.19.2.2 输出量

对比节点输出量如表 186 所示。

表186 对比节点输出量列表

名称	类型	描述
----	----	----



5.19.2.3 配置参数



图302 对比节点属性

对比节点配置参数如图 302 所示，各参数描述如下：

- 节点名称：节点在建模流程图中的名称标注；
- 节点描述：节点描述信息，用于备注节点所实现的功能。

5.19.2.4 使用说明

在数据建模流程中，对比节点位于各模型训练节点的下游。需要注意的是，模型对比节点连接的数据集必须同时包含输入变量与输出变量，且已通过变量设定节点完成了配置。

5.19.3 模型存储

存储节点的功能是将训练好的模型存储至本地磁盘中。

5.19.3.1 输入量

存储节点输入量如表 187 所示。

表187 存储节点输入量列表

名称	类型	描述
model	Model	待存储的模型



5.19.3.2 配置参数



图303 存储节点属性

存储节点配置参数如图 303 所示，各参数描述如下：

- 节点名称：节点在建模流程图中的名称标注；
- 节点描述：节点描述信息，用于备注节点所实现的功能；
- 保存方式：选择保存至模型管理或本地；
- 保存路径：当选择保存方式为本地时，填入保存绝对路径；
- 文件类型：保存的模型文件类型，默认为 dt 格式。

5.19.3.3 使用说明

在数据建模流程中，存储节点位于各模型训练节点的下游。

5.19.4 模型读取

模型读取节点的功能是将资源树中模型管理存储的模型读取至流程中。

5.19.4.1 输出量

模型读取节点输出量如表 188 所示。

表188 模型读取节点输出量列表

名称	类型	描述
model	Model	从资源树模型管理中读取的模型



5.19.4.2 配置参数

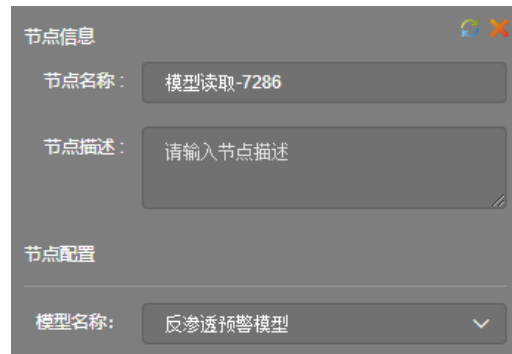


图304 模型读取节点属性

模型读取节点配置参数如图 304 所示，各参数描述如下：

- 节点名称：节点在建模流程图中的名称标注；
- 节点描述：节点描述信息，用于备注节点所实现的功能；
- 模型名称：从资源树中读取的模型名称。

5.19.4.3 使用说明

在数据建模流程中，读取节点一般作为流程中的一个起始点，用户读取模型进行快速评估或模型对比。

5.19.5 模型计算

模型计算节点接收一个模型，支持用户查看该模型的基础信息，并利用该模型对给定的输入变量进行计算，给出输出变量值的计算结果。

5.19.5.1 输入量

模型计算节点输入量如表 189 所示。

表189 模型计算节点输入量列表

名称	类型	描述
model	Model	训练好的模型
data	DTTable	待计算的数据

5.19.5.2 输出量

模型计算节点输出量如表 190 所示。

表190 模型计算节点输出量列表



名称	类型	描述
predictions	DTTable	模型计算后的结果数据

5.19.5.3 配置参数



图305 模型计算节点属性

模型计算节点配置参数如图 305 所示，各参数描述如下：

- 节点名称：节点在建模流程图中的名称标注；
- 节点描述：节点描述信息，用于备注节点所实现的功能。

5.19.5.4 使用说明

在数据建模流程中，模型计算节点一般位于模型训练节点的下游，基于训练好的模型对传入的数据进行计算。

5.19.6 批量训练

批量训练节点支持使用交叉验证的方式使用同样的训练和测试数据集对多个模型进行训练，在提供完成模型的同时，对多个模型的性能给出评估。

5.19.6.1 输入量

批量训练节点输入量如表 191 所示。

表191 批量训练节点输入量列表

名称	类型	描述
meta_models	MetaModel	待训练的多个模型的超参数列表
data	DTTable	训练和测试使用的数据集

5.19.6.2 输出量

批量训练节点输出量如表 192 所示。

表192 批量训练节点输出量列表



名称	类型	描述
result		模型对比结果

5.19.6.3 配置参数



图306 批量训练节点属性

批量训练节点配置参数如图 306 所示，各参数描述如下：

- 节点名称：节点在建模流程图中的名称标注；
- 节点描述：节点描述信息，用于备注节点所实现的功能；
- 测试集占比：测试集占数据总集的比例；
- 随机种子：进行数据划分和交叉验证时的随机种子值，要求是正整数；
- 交叉验证次数：进行交叉验证的次数。

5.19.6.4 使用说明

在数据建模流程中，批量训练节点一般位于多个模型训练节点的下游，结果可以直观的体现当前场景下各类模型的效果。



5.20 时序模型操作

5.20.1 模型预测

模型预测节点接收一个时序模型，支持用户查看该模型的基础信息，并利用该模型对预测变量的未来值进行计算，给出预测结果。

5.20.1.1 输入量

模型预测节点输入量如表 193 所示。

表193 模型预测节点输入量列表

名称	类型	描述
tsmodel	TS-Model	训练好的时序模型

5.20.1.2 输出量

模型预测节点输出量如表 194 所示。

表194 模型预测节点输出量列表

名称	类型	描述
predictions	TS-DTTable	模型预测后的结果数据

5.20.1.3 配置参数

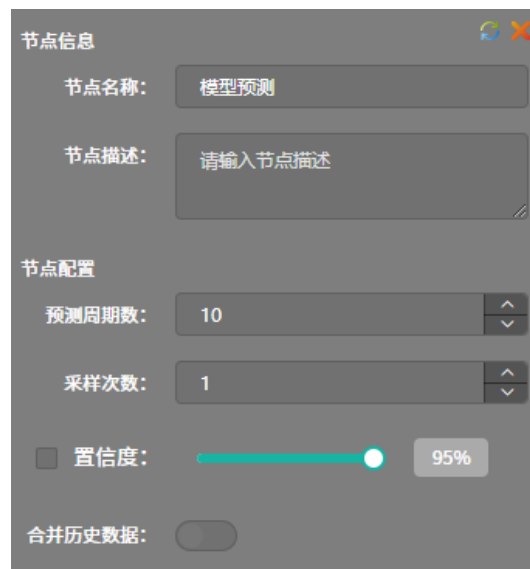


图307 模型预测节点属性

模型预测节点配置参数如图 307 所示，各参数描述如下：

- 节点名称：节点在建模流程图中的名称标注；



- 节点描述：节点描述信息，用于备注节点所实现的功能；
- 预测周期数：从训练集的结束时刻开始，向后预测的时间点个数；
- 采样次数：从概率模型中对预测进行采样的次数（对于确定性模型，应设置为 1）；
- 置信度：开启后，将根据设置的置信度，在输出预测值的基础上同时输出置信区间，包括置信下限（LCL）与置信上限（UCL）；需要注意的是，开启置信区间功能仅对支持的算法有效，详见本手册第 5.18 节；
- 合并历史数据：开启后，在输出数据集中将保留训练数据。

5.20.1.4 使用说明

在数据建模流程中，模型预测节点一般位于时序预测训练算法节点的下游，基于训练好的时序模型对未来值进行预测。

5.20.2 时序模型更新

时序模型更新节点同时接收一个时序模型以及一个时序数据集，支持用户基于最新的数据对原有模型进行修正更新。

5.20.2.1 输入量

时序模型更新节点输入量如表 195 所示。

表195 时序模型更新节点输入量列表

名称	类型	描述
in_tsmodel	TS-Model	待更新的时序模型
data	TS-DTTable	最新的时序数据集

5.20.2.2 输出量

时序模型更新节点输出量如表 196 所示。

表196 时序模型更新节点输出量列表

名称	类型	描述
tsmodel	TS-Model	更新后的时序模型



5.20.2.3 配置参数

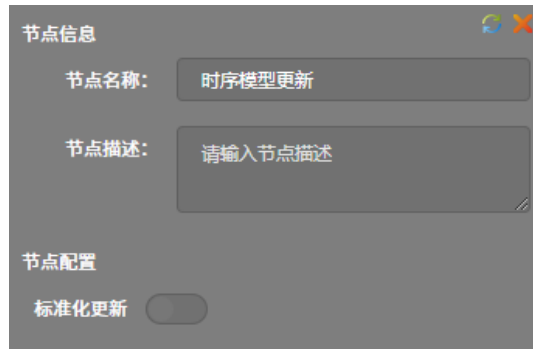


图308 时序模型更新节点属性

时序模型更新节点配置参数如图 308 所示，各参数描述如下：

- 节点名称：节点在建模流程图中的名称标注；
- 节点描述：节点描述信息，用于备注节点所实现的功能；
- 标准化更新：是否更新预测变量及外源变量标准化的类。

5.20.2.4 使用说明

在数据建模流程中，时序模型更新节点一般位于时序预测训练算法节点以及时序预处理的下游，用以更新时序模型。

需要注意的是，该节点的 2 个输入量，即待更新的时序模型 α 与最新的时序数据集 B 之间需保证严格的对应关系。称原时序模型 α 对应的训练集为时序数据集 A 。需保证数据集 A 和 B 二者的时间间隔一致、预测变量相同，且 A 的结束时刻与 B 的起始时刻之间的间隔同样为单个时间步长。

5.20.3 时序模型读取

时序模型读取节点的功能是将资源树中模型管理存储的时序模型读取至流程中。

5.20.3.1 输出量

时序模型读取节点输出量如表 197 所示。

表197 时序模型读取节点输出量列表

名称	类型	描述
tsmodel	TS-Model	从资源树模型管理中读取的时序模型



5.20.3.2 配置参数



图309 时序模型读取节点属性

时序模型读取节点配置参数如图 309 所示，各参数描述如下：

- 节点名称：节点在建模流程图中的名称标注；
- 节点描述：节点描述信息，用于备注节点所实现的功能；
- 模型名称：从资源树中读取的时序模型名称。

5.20.3.3 使用说明

在数据建模流程中，时序模型读取节点一般作为流程中的一个起始点，用户读取时序模型进行快速预测或模型更新。

6 应用案例

6.1 简单案例 1——智能数据挖掘

6.1.1 案例描述

数据挖掘用于揭示数据背后存在的规律。在科学计算、工程设计、生产运维、商务管理等多个领域，从业人员会获得大量的数据，通过数据挖掘将其中所隐含的信息显式地提取出来，可以用于创造更大的价值。在本节中，将介绍包括数据准备、数据清理、模型训练等典型过程在内的精简案例，展示 DTEmpower 在数据挖掘方面的应用。

该案例以基于 Styblinski-Tang 函数采样得到的 299 个算例和 1 个结果异常的算例作为原始数据，部分算例如图 310 所示，其中被红色标注的算例为结果异常的算例。案例原始数据见本手册附件“01-简单案例-智能数据挖掘.csv”。



index	x0	x1	x2	x3	x4	y
1	3.07319	0.691007	-0.92817	-4.30833	1.974288	-39.92
2	2.220556	3.663823	4.755215	3.558033	-4.88286	126.259
3	2.299906	-3.2837	0.210366	-4.45662	-3.00003	-70.549
4	2.936977	-2.76075	-1.54648	4.280813	2.044144	-71.0331
5	-3.35306	1.214784	0.772286	-2.62107	4.34214	-45.8991
6	0.356328	0.8991	2.30122	-1.88055	-1.01779	-63.768
7	-3.13807	4.443724	2.395508	-0.09541	-2.72585	-52.588
8	-4.41971	-0.65583	-1.88204	1.963435	-1.22248	-40.7608
9	-4.75321	-4.3275	1.793928	-0.46303	0.365792	58.75697
10	4.903389	-2.83103	1.630782	-2.36678	-4.79349	90.02777
11	-1.79983	-1.16536	0.883171	3.310485	1.289818	-70.4062
12	-2.26458	2.980468	-3.14364	4.527917	1.874883	-55.6899
13	4.473706	2.308558	-2.46058	-2.86688	0.182007	-46.5333
14	-2.9253	-0.75315	-1.2583	-0.36425	-2.22371	-95.3887
15	-3.50552	3.681261	-3.37507	1.155596	-3.7618	-903.904
16	3.638556	-3.82468	0.173791	-3.67932	2.168597	-75.8907
17	0.654213	-3.1672	-3.55152	-0.11944	-1.44387	-89.1474
18	2.653253	2.486636	4.037197	-4.16578	0.521925	-36.2263
19	4.619364	-2.07852	-2.59171	-3.99706	-4.8357	65.08815
20	1.699165	2.851529	-2.1827	0.864102	-4.36045	-57.8319

图310 案例 1 部分原始数据

案例中 Styblinski-Tang 函数包含五个输入变量 x_0 、 x_1 、 x_2 、 x_3 、 x_4 ，一个输出变量 y ，其中任意两个输入变量组成的函数曲面(其它输入变量值等于零时)如图 311 所示。

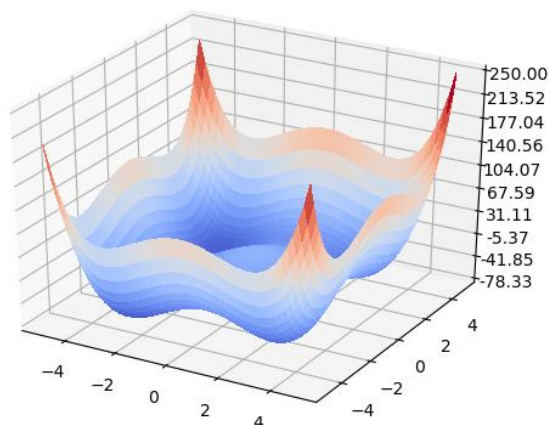


图311 Styblinski-Tang 函数曲面



针对该案例，利用 DTEmpower 提供的智能数据清理算法 AIOD 和智能代理模型训练算法 AI-Agent 进行智能数据挖掘。

6.1.2 操作流程

6.1.2.1 添加数据集

在顶部菜单栏“数据处理”中，单击“上传按钮”，即跳出数据集文件上传窗口。点击窗口中“选择文件”按钮，即可选择本地磁盘中的数据文件加载至当前项目中，如图 312 所示。

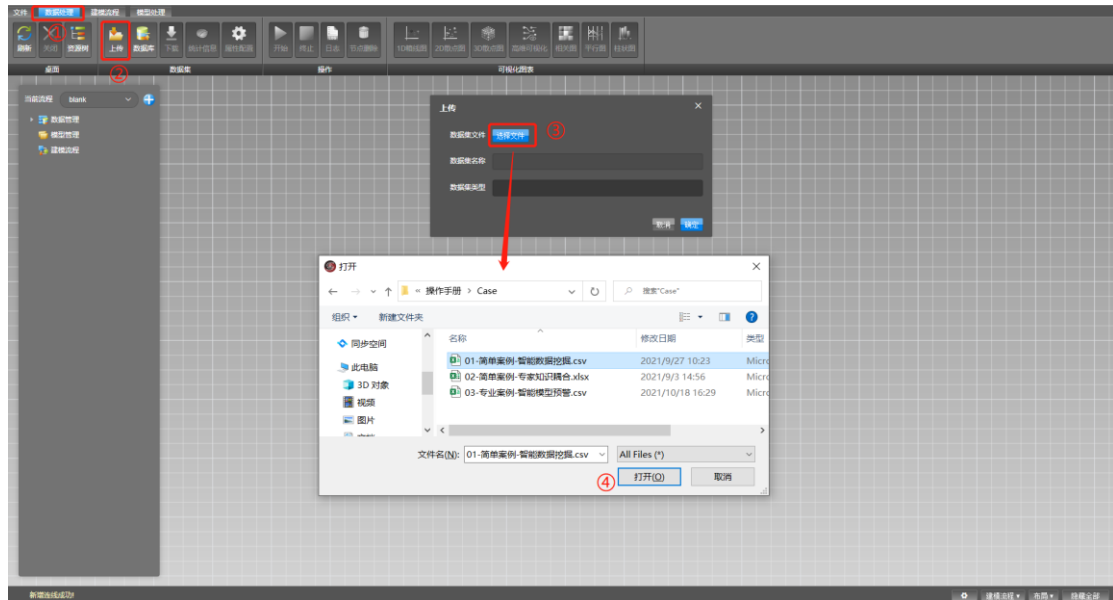


图312 上传 Styblinski-Tang 原始数据

数据集添加完成后即出现在左侧资源菜单中如图 313 所示。

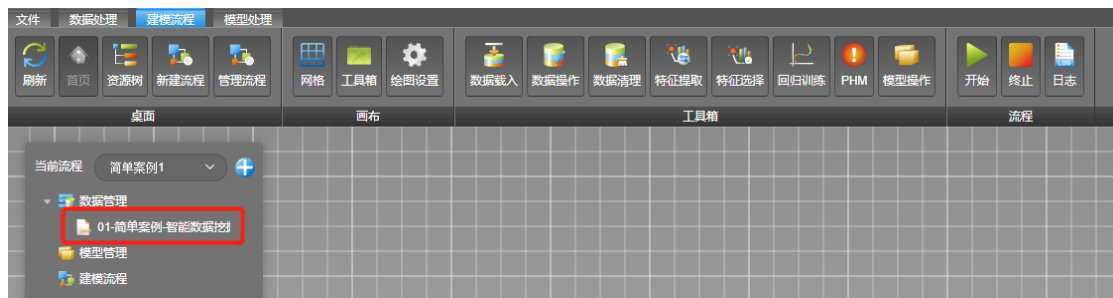


图313 数据集上传结果



6.1.2.2 加入流程

在已添加的数据集上单击鼠标右键，在右键快捷菜单中选择“加入流程”菜单，原始数据的载入即以“数据读取”节点的形式加入至建模流程中。在建模画布与左侧资源树中均可以观察到新增的节点，如图 314 所示。

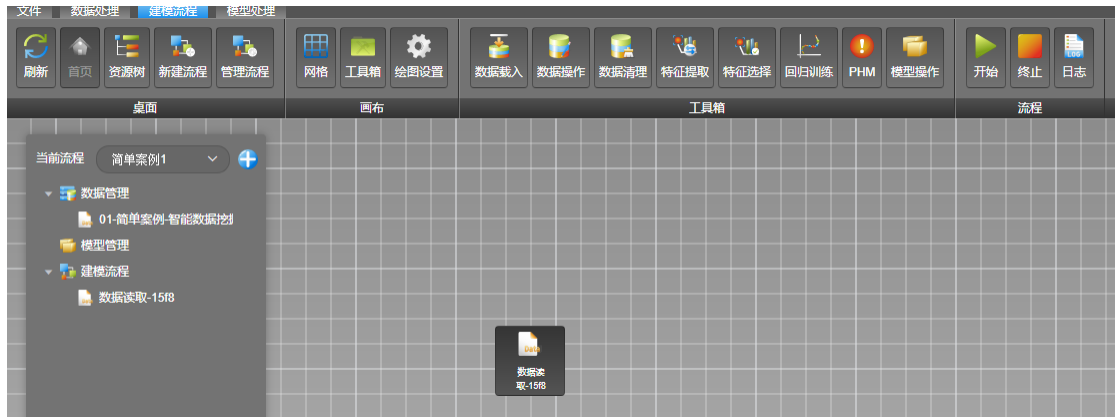


图314 加入流程

6.1.2.3 变量设定

在建模流程中，原始数据载入后，需要依托“数据操作-变量设定”节点将变量指定为输入变量和输出变量两类，后续的数据清洗、特征工程和模型训练都将依托于输入和输出变量进行。

在顶部菜单栏中单击工具箱中的“数据操作”，即可展开右侧工具箱。选择其中的“变量设定”控件并拖拽至画布中，即完成了新节点的添加，如图 315 所示。

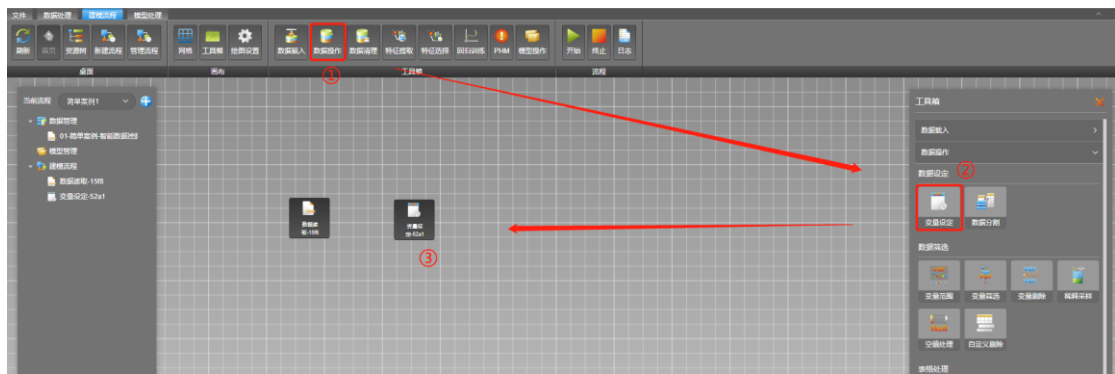


图315 添加变量设定节点

此时可以单击添加至建模流程中的变量设定节点，在节点属性弹窗中修改节点名、节点描述，方便查看和理解，如图 316 所示。需要注意的是，该操作不是必须的。



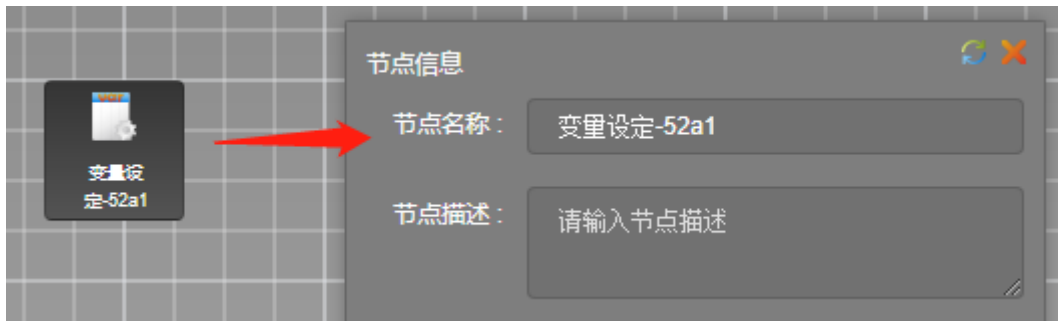


图316 修改节点名称及描述

数据读取节点读取的原始数据将传递至变量设定节点，进行输入和输出变量的设定，因此需要连接数据读取节点和变量设定节点，如图 317 所示。

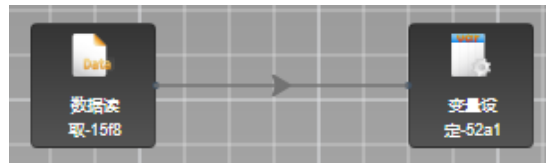


图317 连接数据载入和变量设定节点

数据读取节点和变量设定节点连接完成后，会自动弹出节点间输入输出量的关联列表，系统会默认对类型匹配的输入输出量进行自动关联，用户可以根据自身需求进行增删改操作，如图 318 所示。



图318 连线配置

6.1.2.4 异常点清洗

DTEmpower 提供自研智能数据清理算法 AIOD，节点执行完成后，会为每条算例提供一个异常风险评分，分值越高该算例为异常点的可能性越高。

将工具箱切换为数据清理，然后将 AIOD 节点加入至建模流程，如图 319 所示。



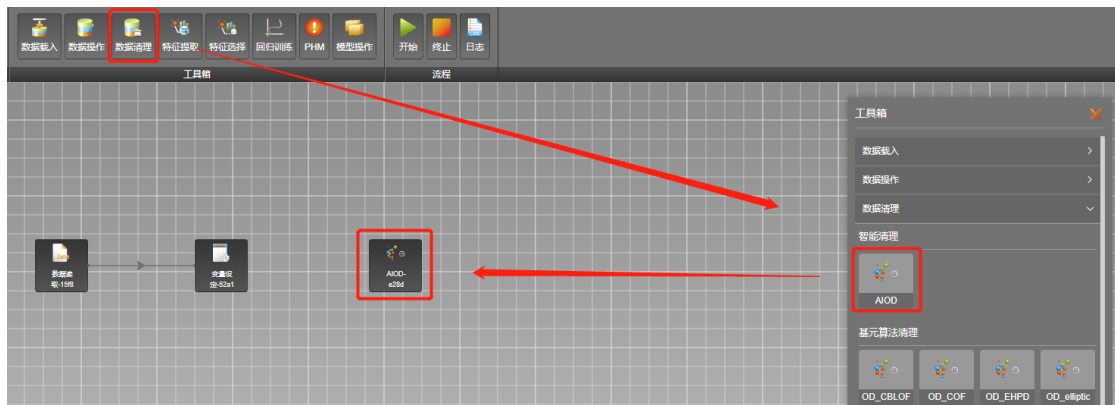


图319 添加 AIOD 节点

同变量设定节点一样，用户可通过单击 AIOD 节点的方式修改节点名称和描述，具体操作上文已经进行详细描述，此处不再赘述。

经变量设定节点指定输入输出变量的数据将传递至 AIOD 节点，进行数据清洗，因此需要连接变量设定节点和 AIOD 节点，如图 320 所示。



图320 连接变量设定节点和 AIOD 节点

如需调整节点的位置与距离以保持画布整洁，可以同时按住键盘“Ctrl”键并用鼠标进行框选，在弹出菜单中选择对齐方式等，如图 321 所示。

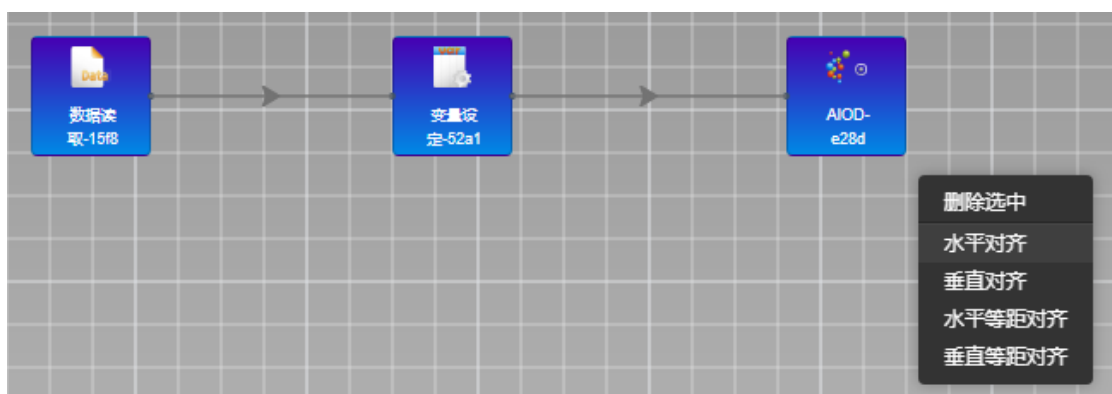


图321 节点框选与对齐

6.1.2.5 智能模型训练——AI Agent

DTEmpower 提供自研智能模型训练算法 AI Agent，内嵌集成学习、数据分层分类及超参自学习模块的 AI Agent 无需用户具有算法基础或调参经验。



将工具箱切换为回归训练，然后将 AIAgent 节点加入至建模流程，如图 322 所示。

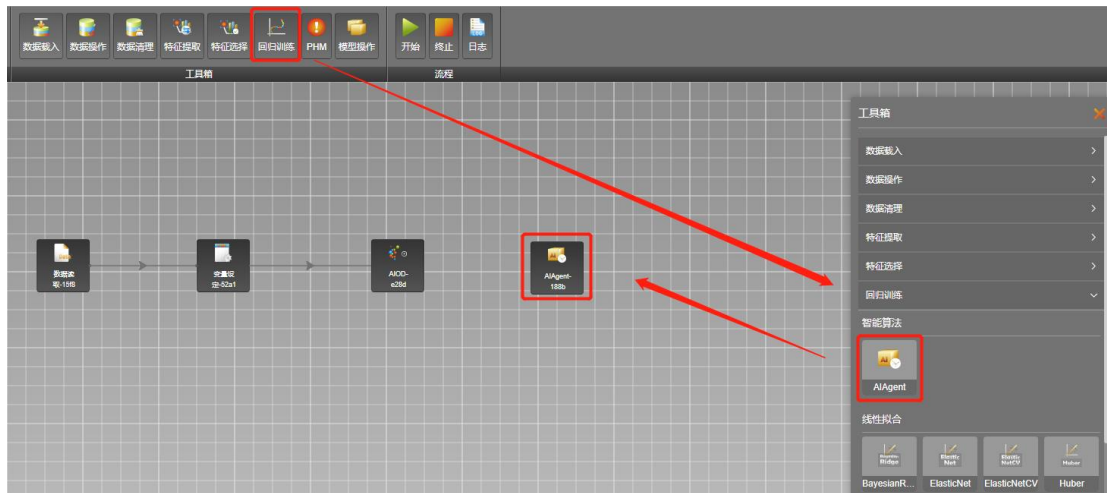


图322 添加 AIAgent 节点

点击画布上的 AIAgent 节点，右侧将弹出节点配置面板。开启其中的“小数据集优化”选项，如图 323 所示。该选项适合在样本量在 1000 行上下的数据集开启，将耗费更多的训练时长，但模型精度也相对更高。



图323 AIAgent 配置面板

采用同样的方式可修改节点的名称，并完成节点间的连接，如图 324 所示。

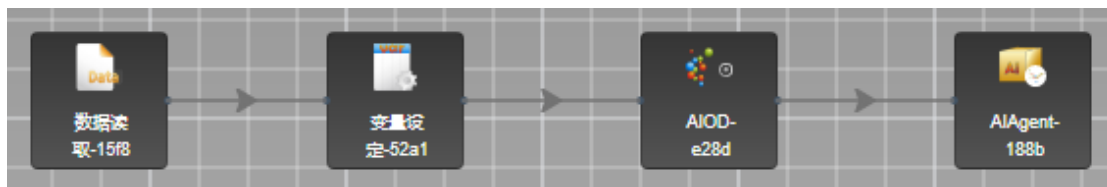


图324 连接 AIOD 节点和 AIAgent 节点



6.1.2.6 执行流程与交互式配置

建模流程搭建完成后，单击顶部菜单栏“开始”按钮后，建模流程中的节点即开始依次执行，其中执行至变量设定和 AIOD 节点时流程执行会自动暂定，用户在交互式配置弹窗中配置完成后才会继续向下执行，直至建模流程执行完成。

执行至变量设定节点时，自动弹出的交互式配置弹窗如图 325 所示。

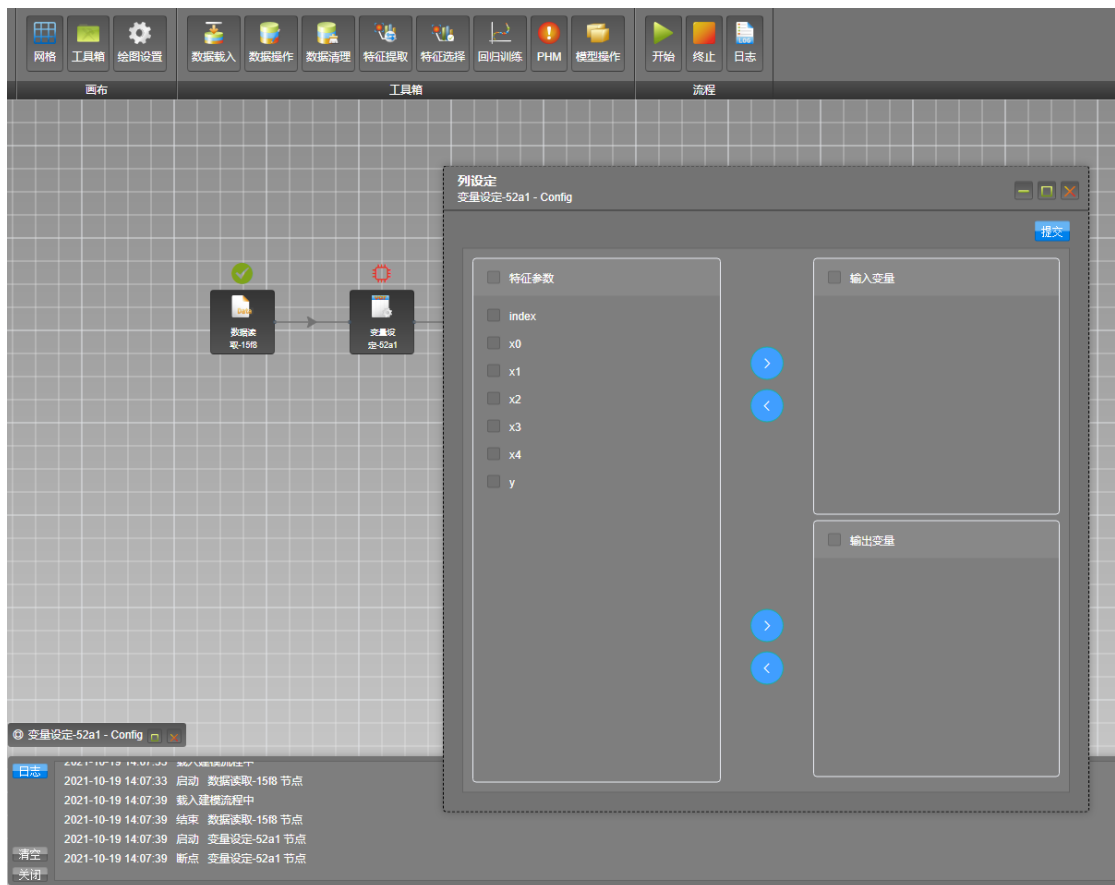


图325 变量设定交互弹窗

如前文所述，变量设定交互式弹窗左侧为从原始数据文件表头中读取到的变量名称，右侧上栏为输入变量，右侧下栏为输出变量。用户在左侧勾选变量，然后单击表格间的移动按钮，即可实现输入和输出变量的设定，分别如图 326、图 327 所示。在勾选变量时，支持 Shift 快捷键批量选择。



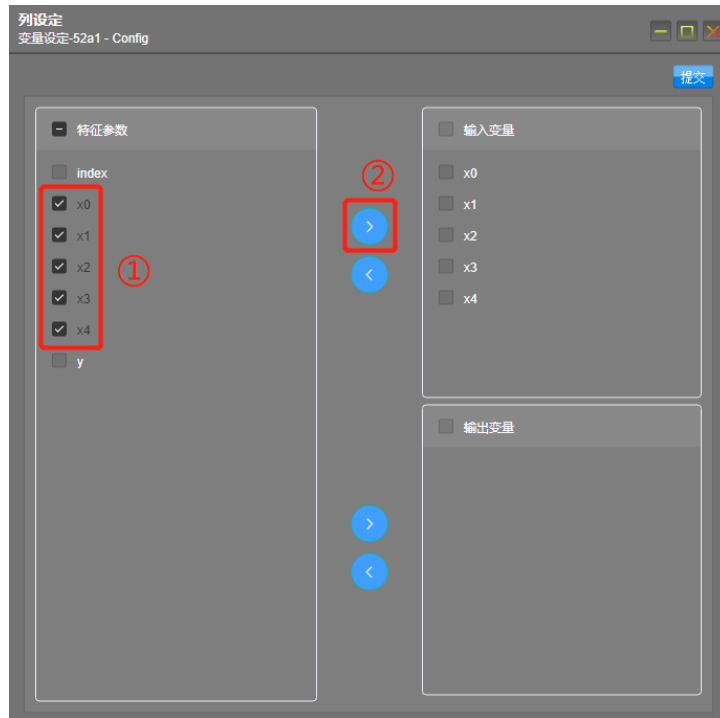


图326 输入变量设定



图327 输出变量设定

输入输出变量设定完成后，单击交互式弹窗右上角的提交按钮即可继续向下执行，如图 328 所示。



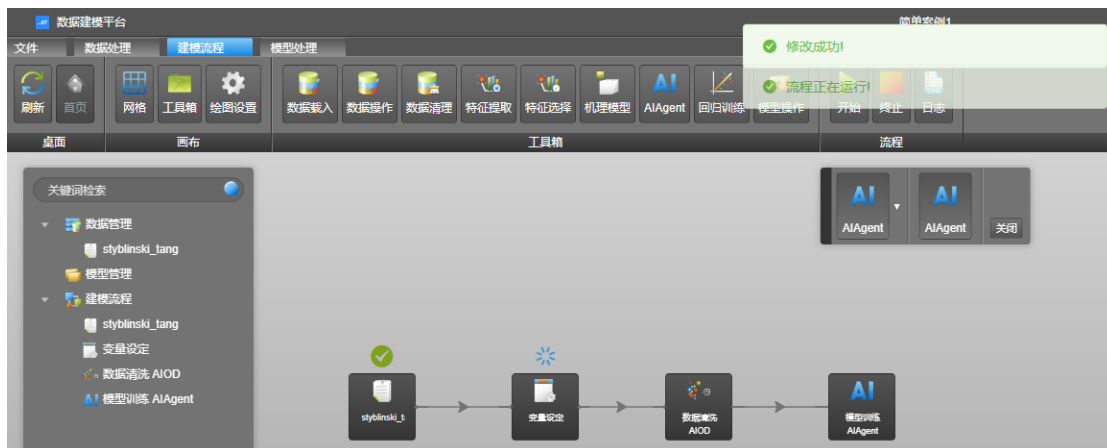


图328 提交变量设定

流程执行至 AIOD 节点后，会自动弹出数据清理的交互式弹窗，用户在该弹窗中可根据右侧列表中各算例给定的异常风险值辅以左侧的可视化图表，选择异常点。需要注意的是，右侧的算例列表是按异常风险评分由高到低排序的。左侧的可视化图表包含三种形式：柱状图、箱线图以及散点图，如图 329 所示。

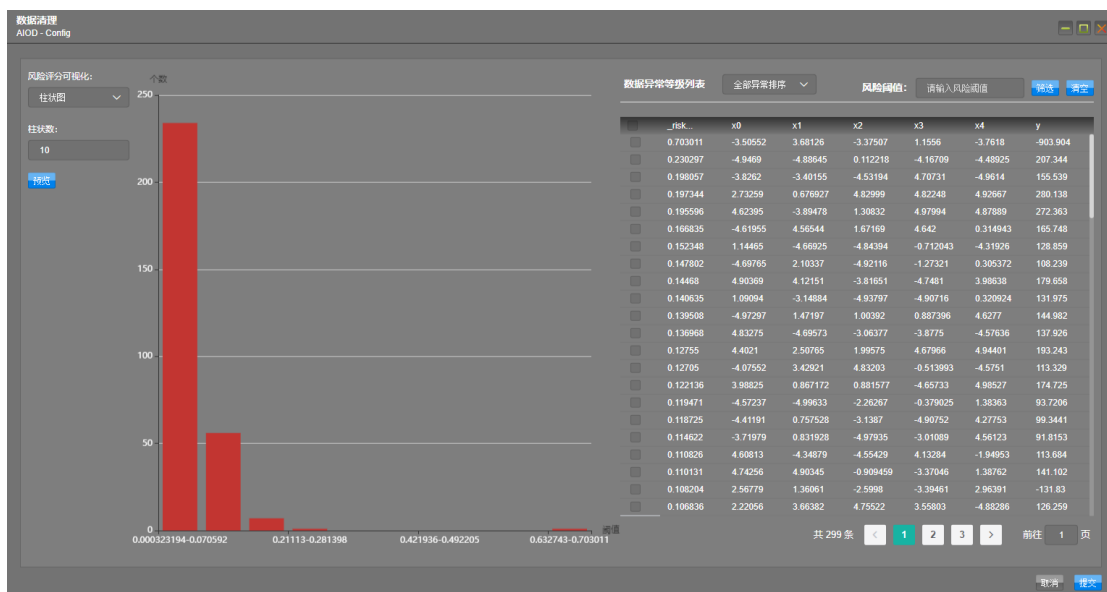


图329 AIOD 节点交互式弹窗

通常而言，异常风险评分明显高于其他算例的算例是异常点的概率更高。在右侧列表中勾选或左侧散点图中双击算例点均可将对应算例指定为异常点；在右侧列表中取消勾选或在左侧散点图中双击已经被指定为异常点的算例，即可将取消对应算例的异常点指定。点击左侧“预览”按钮可以查看剔除异常点后的新可视化显示。



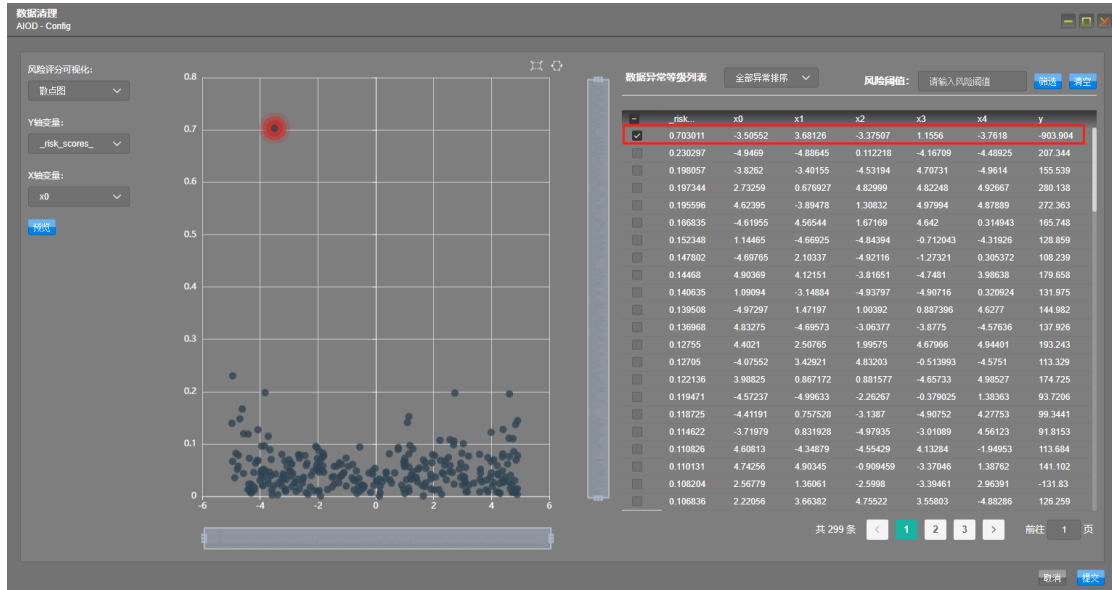


图330 异常点配置

如图 330 所示，异常风险评分最高的算例即为真实的异常点，其异常风险评分在 0.70 以上，而其它算例异常风险评分最高为 0.23。因此，将其指定为异常点后单击交互式弹窗右上角的提交按钮即可继续向下执行，直至整个流程执行完成，如图 331 所示。

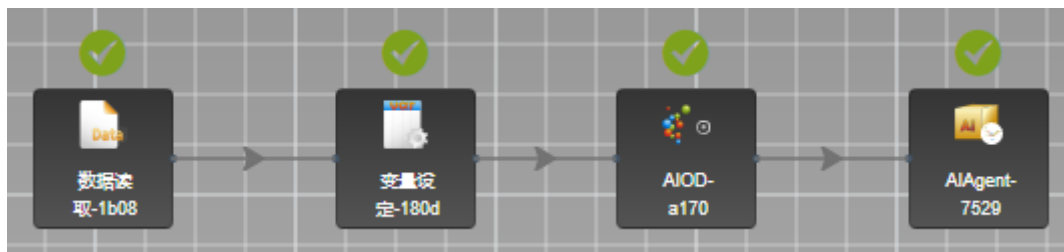


图331 流程执行完成

6.1.3 效果对比

建模流程执行完成后，展开资源树中 AI Agent 节点，单击 model 选项，依次选择输出变量、勾选响应面自变量、并配置其它输入变量的变量值后，单击开始绘图按钮后，右侧即显示训练得到的模型响应面信息，如图 332 所示。



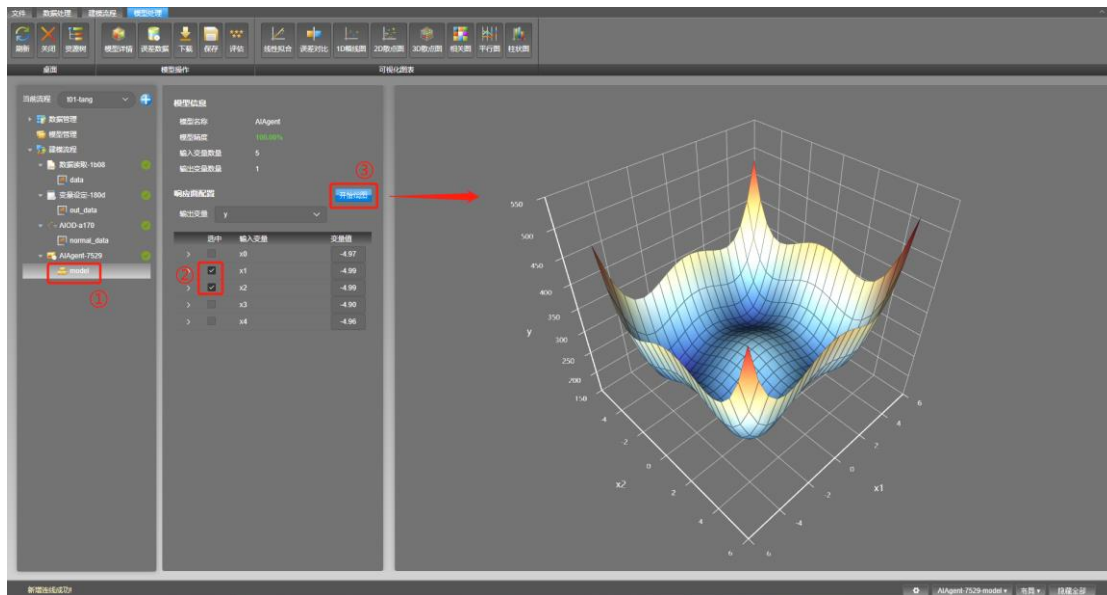


图332 模型信息查看

为了验证 AI Agent 的优越性，本案例同时新建了对比工程，利用集成学习算法 Adaboost 节点代替 AI Agent 节点，在其它配置均相同的情况下，两者响应面对比如图 333 所示。

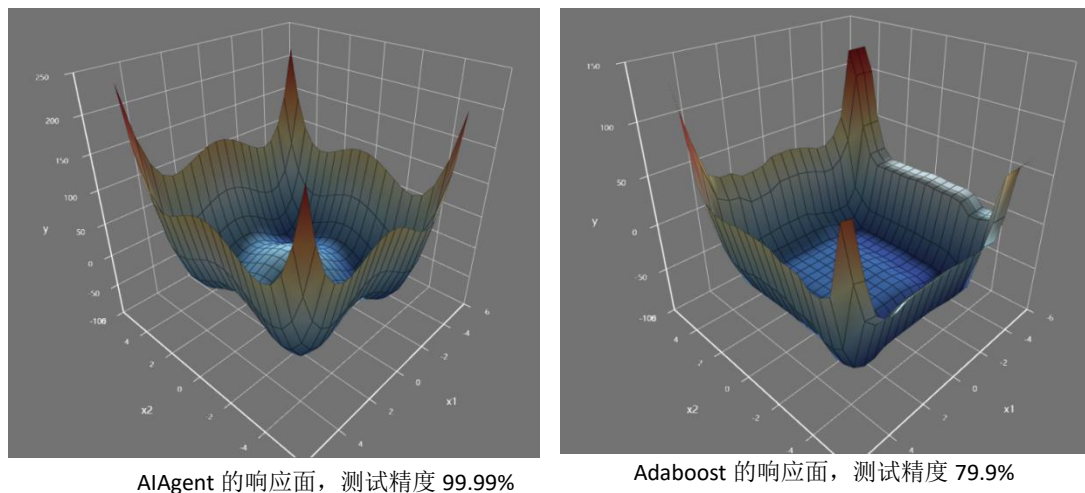


图333 训练效果对比

由图 333 可知，AI Agent 响应面结果与真实函数空间趋势几乎完全一致，而 Adaboost 的响应面绝大部分区域的细节信息丢失。



6.2 简单案例 2——专家知识耦合

6.2.1 案例描述

在基于数据驱动的建模流程中，耦合专家知识可以大大增加模型的可解释性，且通常能够增强模型的泛化能力。在复杂工业场景应用中，专家知识通常是不可或缺的。在本案例中，将介绍在 DTEmpower 的建模流程中如何耦合专家知识。

本案例基于如式 1 所示的函数进行原始数据采样，其中 x_1 和 x_2 作为输入变量， y 作为输出变量。

$$\begin{cases} x_3 = x_1 * x_2 \\ y = a * (x_3 + b)^2 + c \end{cases} \quad \text{式 1}$$

采样过程中，指定输入变量的范围为 $[0, 1]$ 、 $a=7$ 、 $b=-0.4$ 、 $c=1$ ，并对采样结果中添加了 10% 的噪声，用以模拟实际生产生活中的干扰，随机采样获得的 50 组算例作为本案例的原始数据，如图 334 所示，需要注意的是 x_3 作为中间变量，对采样过程而言是不可知的。案例原始数据见本手册附件“02-简单案例-专家知识耦合.xls”。

index	x1	x2	y
1	0.548814	0.715189	1.060693
2	0.544883	0.423655	1.277827
3	0.437587	0.891773	1.097099
4	0.383442	0.791725	1.121409
5	0.568045	0.925597	1.118635
6	0.087129	0.020218	2.285853
7	0.778157	0.870012	1.687552
8	0.799159	0.461479	1.085401
9	0.118274	0.639921	1.761145
10	0.944669	0.521848	1.104484

图334 案例 2 部分原始数据

针对该案例，利用 DTEmpower 提供的特征提取功能和机理模型拟合算法实现数据与专家知识的耦合。



6.2.2 操作流程

6.2.2.1 添加数据集

在工程树状资源菜单“数据管理”项上单击鼠标右键，在右键快捷菜单中单击上传按钮，在弹出窗口中选择本地磁盘中的数据文件后单击确定即可完成数据集的添加，如图 335 所示。

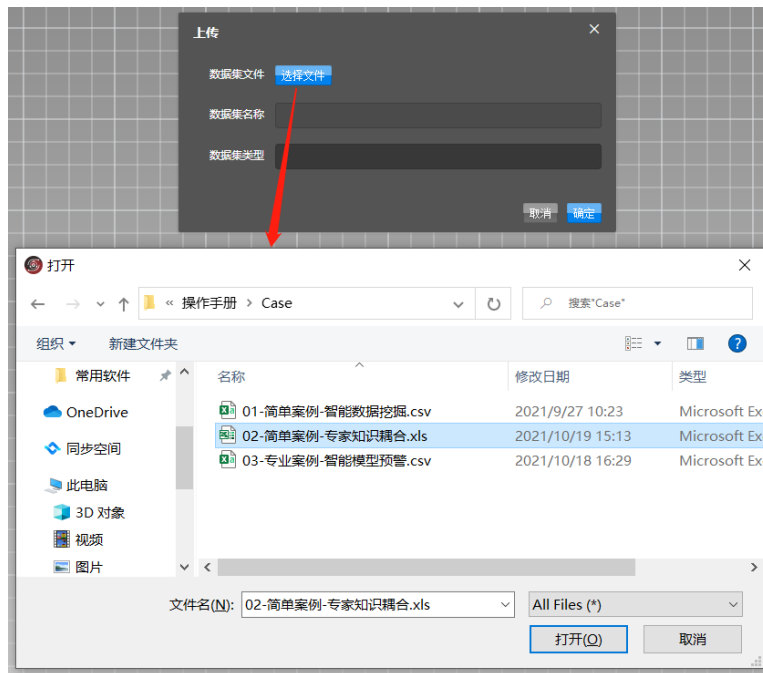


图335 上传案例原始数据

数据集添加完成后即出现在左侧资源菜单中如图 336 所示。

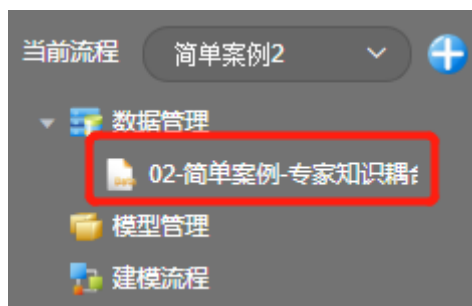


图336 数据集上传结果

6.2.2.2 加入流程

在已添加的数据集上单击鼠标右键，在右键快捷菜单中选择加入流程菜单，原始数据的载入即以数据读取节点的形式加入至建模流程中如图 337 所示。



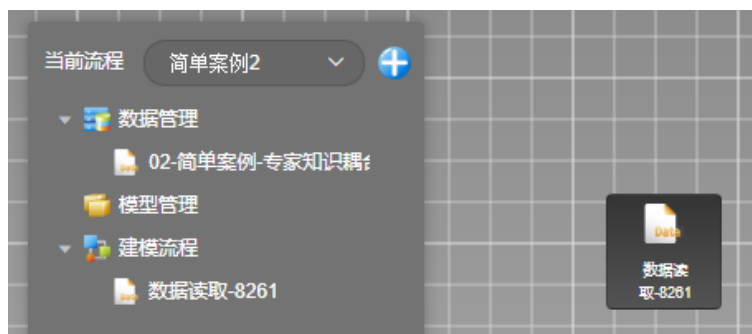


图337 加入流程

6.2.2.3 特征生成

利用特征生成，用户可将积累的工程经验或专家知识加入建模流程中，这种已知的参数间的部分机理关系，能够辅助数据挖掘算法，在有限数据集的条件下，建立精度更高，可解释性和迁移性更好的模型。

此处假设用户已知中间变量 x_3 与输入变量之间的关系，即可使用自定义特征提取节点将其加入建模流程中，具体操作如下。

将工具箱切换为特征生成，然后将表达式节点加入至建模流程，如图 338 所示。

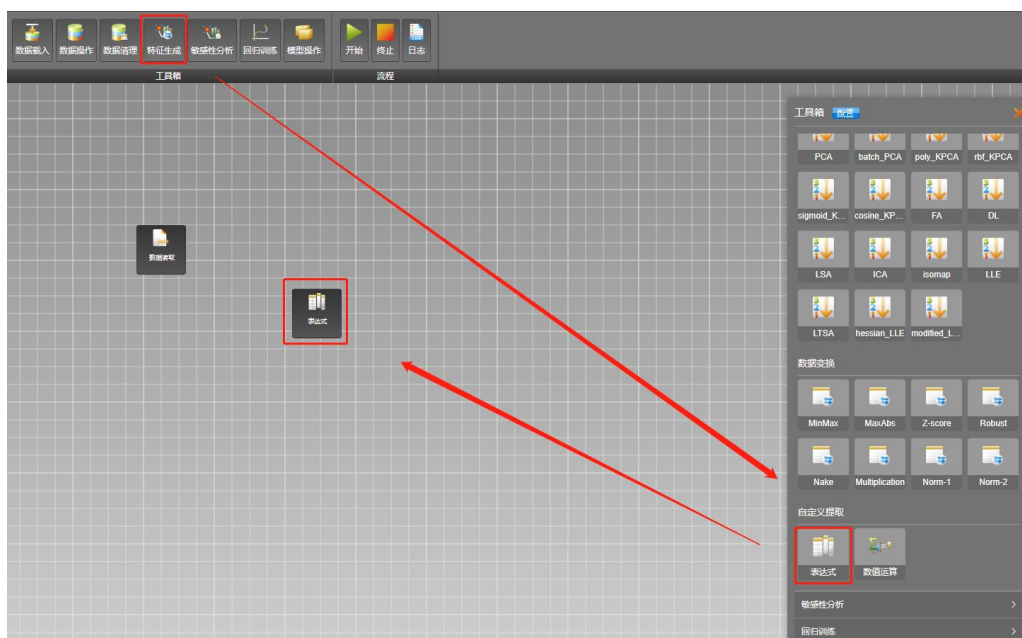


图338 添加表达式节点

同前述操作相同，用户可对添加至建模流程中的自定义提取节点名称和描述进行编辑，此处不再赘述。



数据读取节点读取的原始数据将传递至表达式节点，进行中间变量 x_3 的设定，因此需要连接数据读取节点和表达式节点，如图 339 所示。

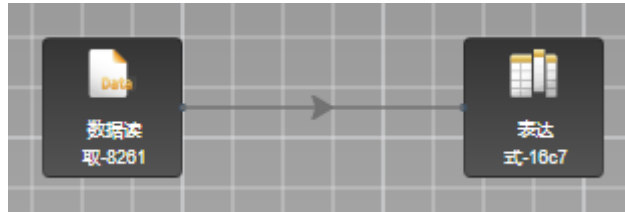


图339 连接数据载入节点和自定义提取节点

6.2.2.4 变量设定

机理模型拟合将依托于输入和输出变量进行，因此需要对输入输出变量进行设定。

将工具箱切换为数据操作，然后将变量设定节点加入至建模流程，如图 340 所示。

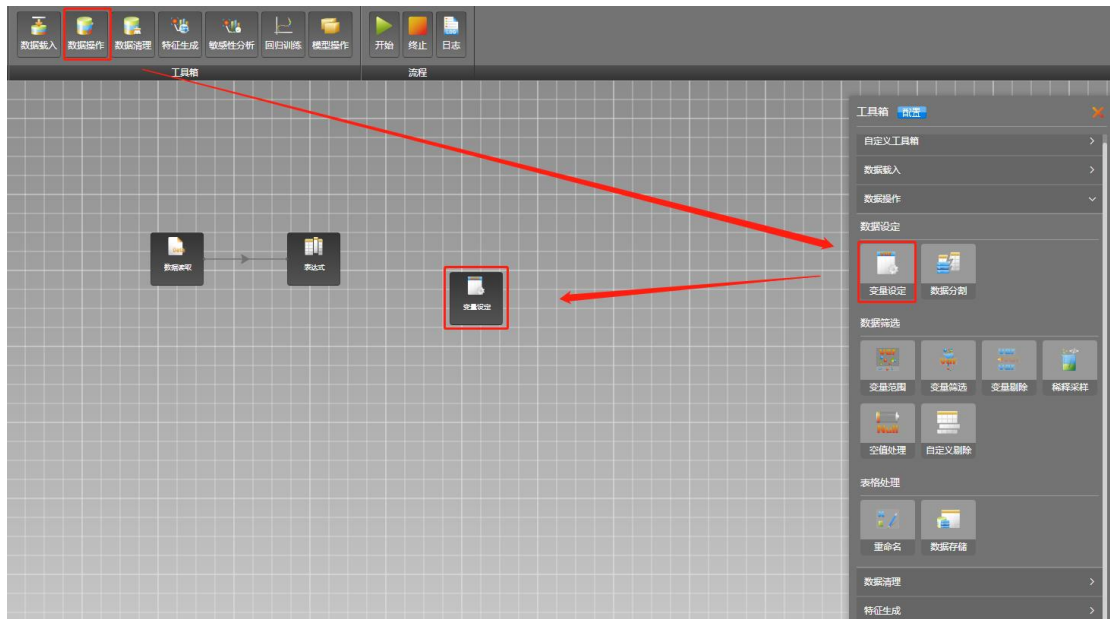


图340 添加变量设定节点

变量设定节点添加至建模流程后，可采用与前文描述相同的方法对其节点名称和节点名称进行编辑，此处不再赘述。

通过表达式节点增加中间变量 x_3 的数据将传递至变量设定节点，进行输入和输出变量的设定，因此需要连接自定义提取节点和变量设定节点，如图 341 所示。



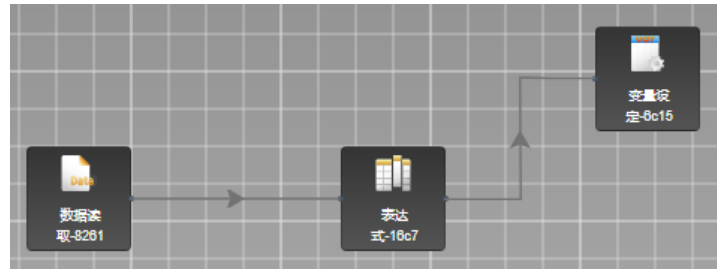


图341 连接表达式和变量设定节点

6.2.2.5 机理模型拟合——显函数

DTEmpower 提供机理模型拟合节点显函数，它能基于训练数据，对机理关系中的未知参数进行求解，这些未知参数往往与工况和外部条件有关，是不确定的。

此处假设用户已知中间变量 x_3 与输出变量之间的关系，即可使用显函数节点将其加入建模流程中，具体操作如下。

将工具箱切换为回归训练，然后将显函数节点加入至建模流程，如图 342 所示。

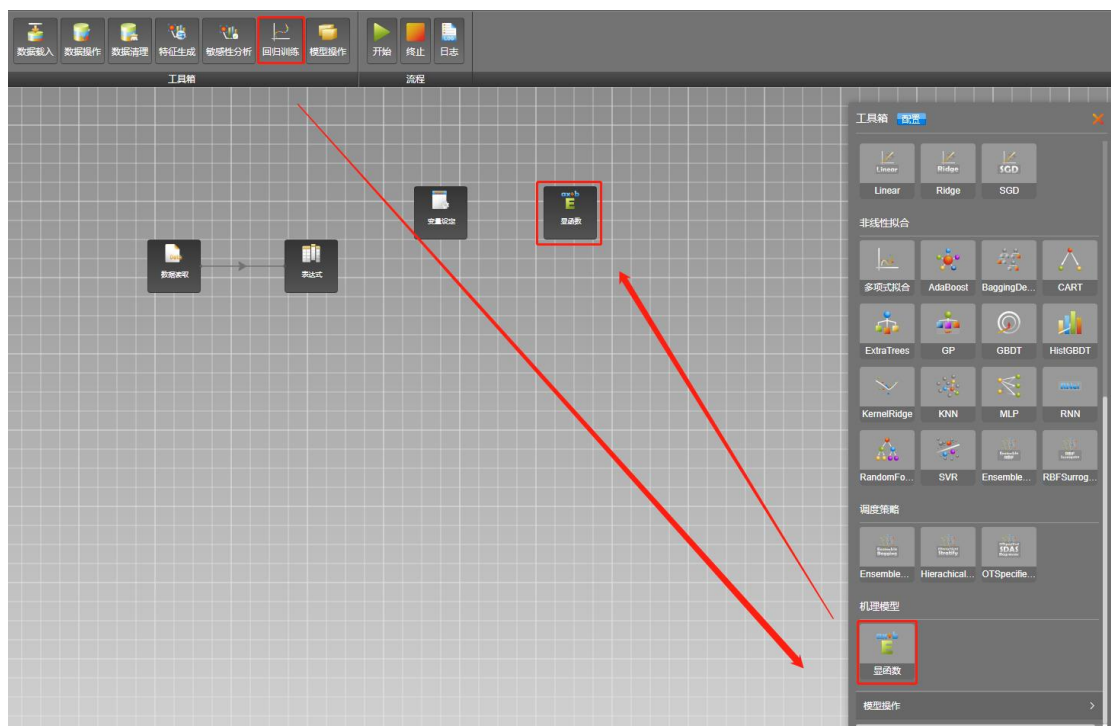


图342 添加显函数节点

采用同样的方式可修改节点的名称，并完成节点间的连接，如图 343 所示。



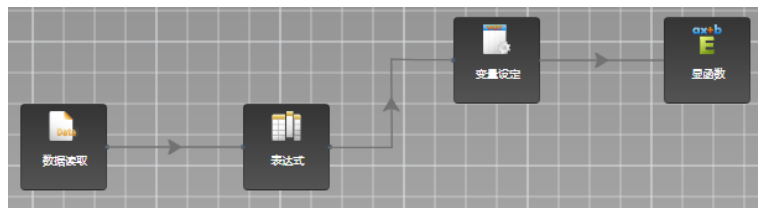


图343 连接变量设定节点和显函数节点

6.2.2.6 模型融合

基于工程经验的特征生成和机理模型拟合需要通过 DTEmpower 的模型聚合功能固化，固化后的模型接收输入变量 x_1 、 x_2 ，在计算得到中间变量 x_3 后，利用机理模型拟合得到的 x_3 到输出变量 y 的映射关系预测输出变量 y 的值。

将工具箱切换为模型操作，然后将聚合节点加入至建模流程，如图 344 所示。

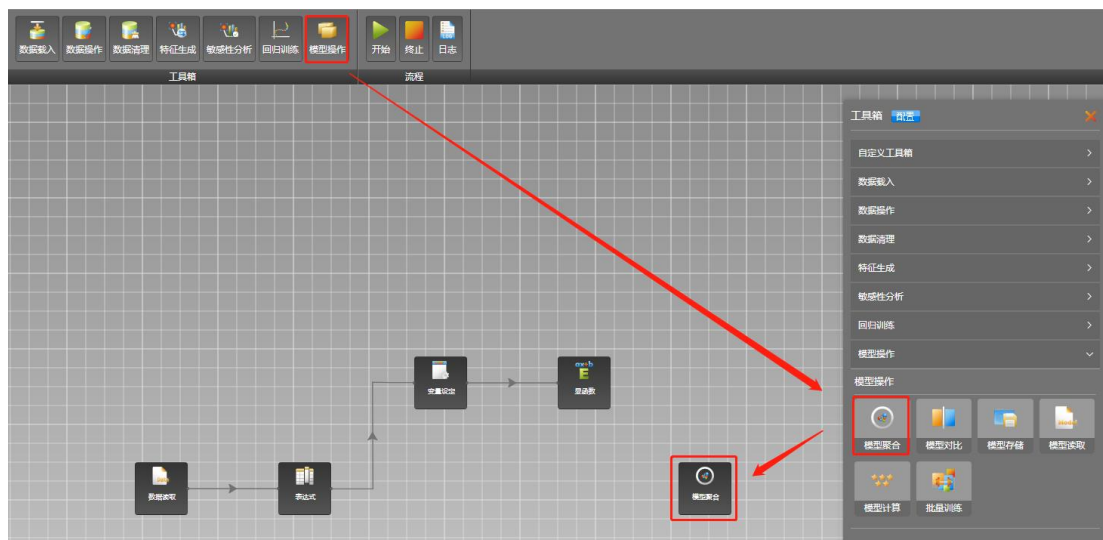


图344 添加聚合节点

采用同样的方式可修改节点的名称，并完成节点间的连接，如图 345 所示。需要注意的是，特征提取规则和机理模型都应被记录和保存，因此表达式节点和显函数节点均需连接至聚合节点。

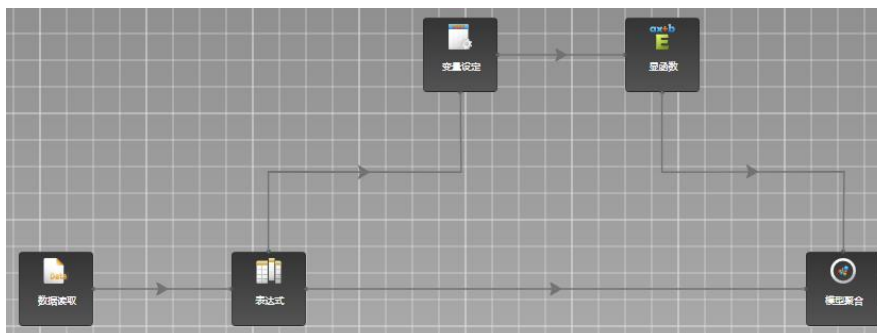


图345 连接聚合节点



6.2.2.7 执行流程与交互式配置

建模流程搭建完成后，单击“开始”按钮后，建模流程中的节点即开始依次执行，其中执行至表达式节点、变量设定节点和显函数节点时流程执行会自动暂定，用户在交互式配置弹窗中配置完成后才会继续向下执行，直至建模流程执行完成。

执行至表达式节点时，自动弹出的交互式配置弹窗如图 346 所示。

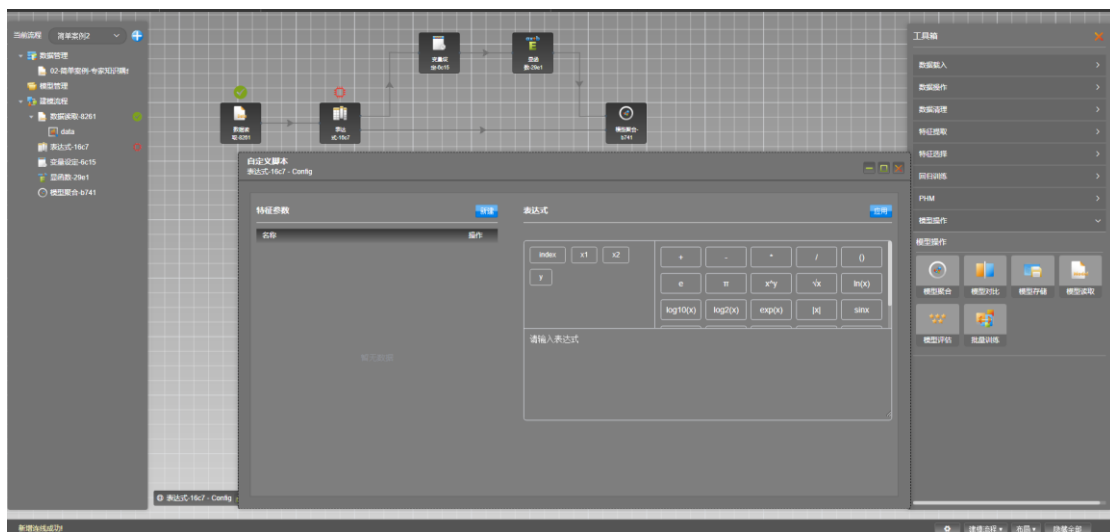


图346 自定义特征提取-表达式节点

左侧为新提取的特征（变量）列表，用户可通过新建按钮添加新的特征（变量），右侧为自定义特征的计算公式，同时提取多个特征时，每个特征计算公式为一行。针对本案例的特征提取配置结果如图 347 所示。



图347 中间变量 x_3 的提取配置

特征提取完成后，单击交互式弹窗右上角的应用按钮即可继续向下执行，直至运行至变量设定节点。

执行至变量设定节点时，自动弹出的交互式配置弹窗如图 348 所示。



图348 变量设定交互弹窗

采用与前述案例相同的操作在变量设定交互式弹窗中对输入输出变量进行设定，对于本案例而言，其设定结果如图 349 所示。



图349 输入输出变量设定结果



输入输出变量设定完成后,单击交互式弹窗右上角的提交按钮即可继续向下执行,直至运行至显函数节点。

执行至显函数节点时,自动弹出的交互式配置弹窗如图 350 所示。

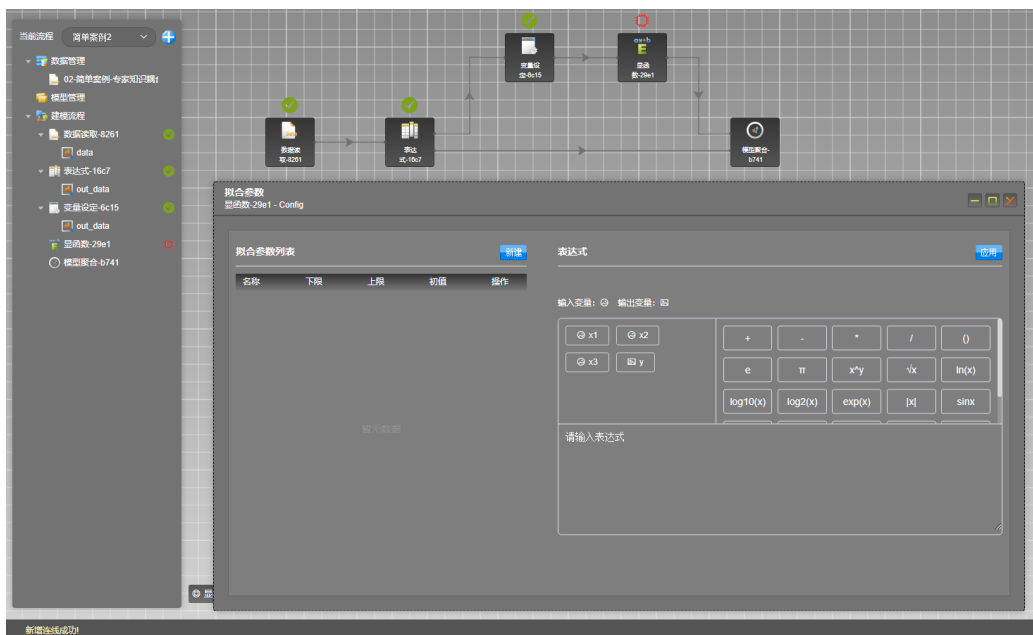


图350 机理模型拟合交互弹窗

其中弹窗左侧为待拟合参数列表,右侧为输出变量的计算公式。本案例中存在 3 个待拟合的参数 a、b、c,其与中间变量 x_3 和输出变量 y 的关系如式 1 所示,因此本案例的显函数拟合配置如图 351 所示。优化算法选择 SLSQP。

$$\begin{cases} x_3 = x_1 * x_2 \\ y = a * (x_3 + b)^2 + c \end{cases}$$

式 1 所示



图351 机理模型拟合配置

需要注意的，对于本案例而言，也可单独利用显函数节点直接拟合输入变量 x_1 、 x_2 和输出变量 y 之间的关系，但直接拟合的缺点是训练得到的模型不包含中间变量 x_3 的信息。

机理模型拟合节点设定完成后，单击交互式弹窗右上角的应用按钮即可继续向下执行，直至运行至建模流程执行完成，如图 352 所示。

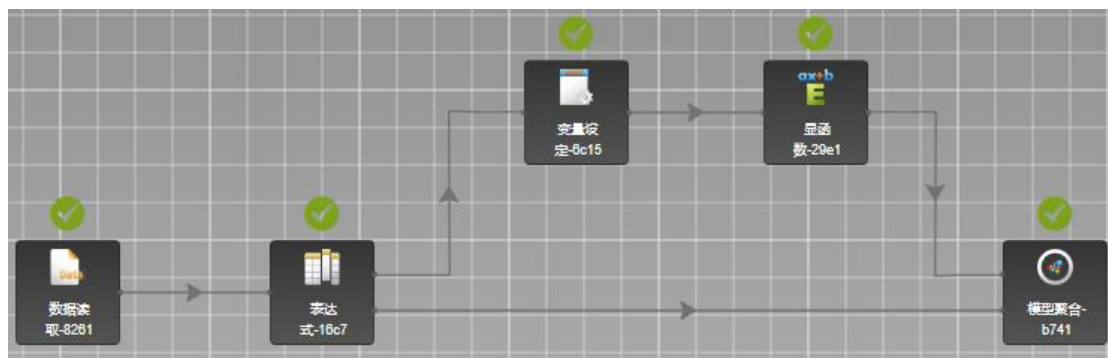


图352 建模流程执行完成

6.2.3 效果对比

建模流程执行完成后，展开资源树中模型聚合节点，单击 model 选项，依次选择输出变量、勾选响应面自变量、并配置其它输入变量的变量值后，单击开始绘图按钮后，右侧即显示训练得到的模型响应面信息，如图 353 所示。需要注意的，对于聚合后的模型而言，其输入变量为 x_1 、 x_2 ，输出变量为 x_3 和 y 。

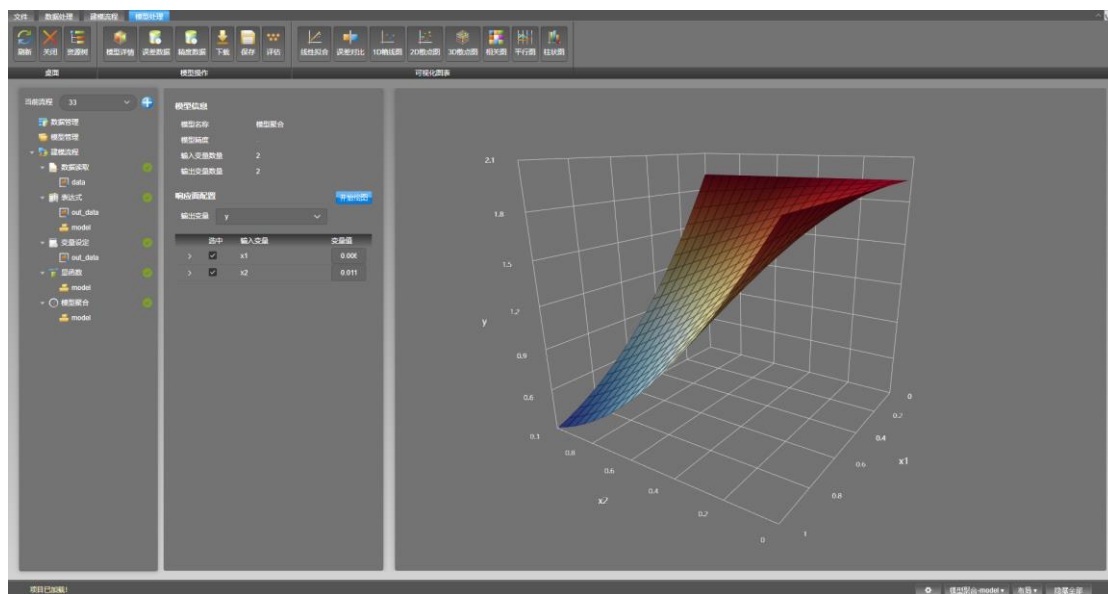


图353 模型信息查看

为了验证专家知识耦合的优越性，本案例同时新建了对比工程，分别利用多项式拟合和集成学习算法 Adaboost 直接拟合输入变量 x_1 、 x_2 和输出变量 y 之间的关系，在其它配置均相同的情况下，三者拟合对比如图 354 所示。

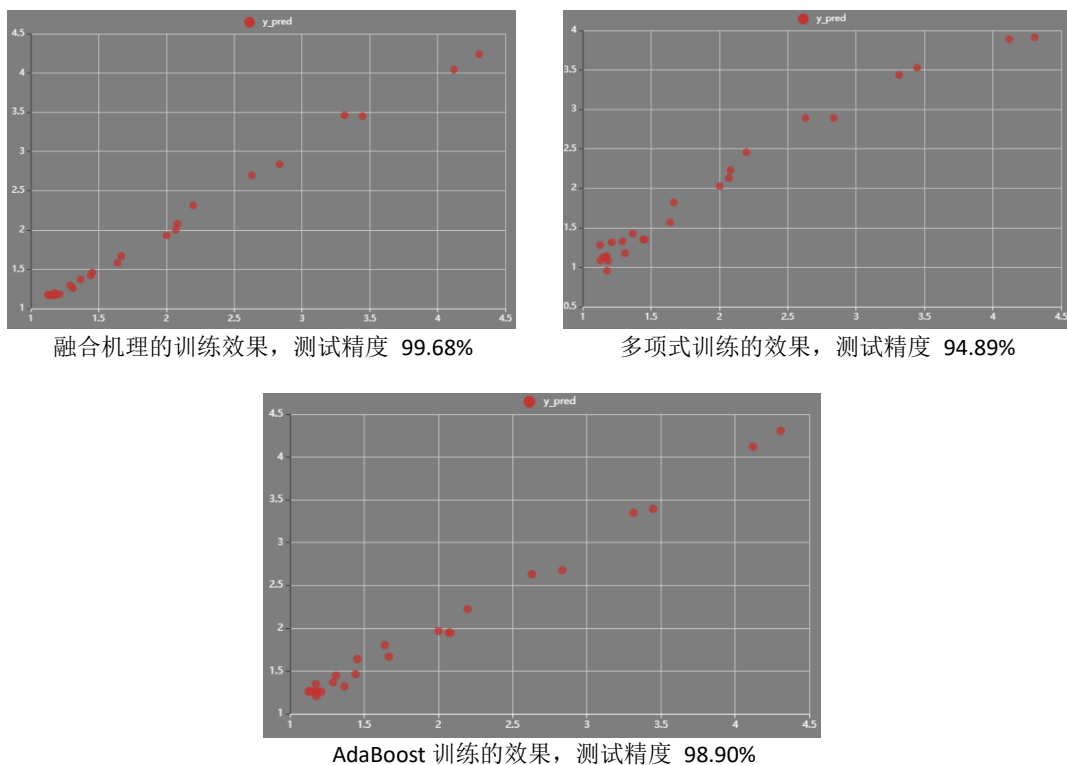


图354 训练效果对比

由图 354 可知，融合了机理的 DT 模型的测试精度为 99.68%，相对于其他对比算法有均所提高，证明了 DTEmpower 在机理融合方面的强大能力。

6.3 简单案例 3——数据前处理与可视化

6.3.1 案例描述

准备合适的数据集是开展建模前的必要准备工作，因此一个高效、易用的数据前处理与可视化工具可以帮助数据分析与数据建模工作事半功倍。在本案例中，将会介绍如何拼接来自不同数据源的样本为同一个数据集，如何剔除无效样本，以及如何通过可视化对数据集形成初步判断。



本案例中准备的数据集包含 2 个数据文件，采集自某电厂前置泵运行数据。两个文件中的变量完全一致，均包含时间、流量、电流以及压力数据。但其中一个数据集采集自设备正常运行阶段，另一个设备采集自设备异常运行阶段。

6.3.2 操作流程

6.3.2.1 上传数据集

在工程树状资源菜单“数据管理”项上单击鼠标右键，在右键快捷菜单中单击上传按钮，在弹出窗口中选择本地磁盘中的数据文件后单击确定即可完成数据集的添加，如图 355 所示。在本案例中，需先后添加 2 个数据集，见本手册附件“03a-简单案例-数据前处理与可视化.xls”和“03b-简单案例-数据前处理与可视化.xls”。

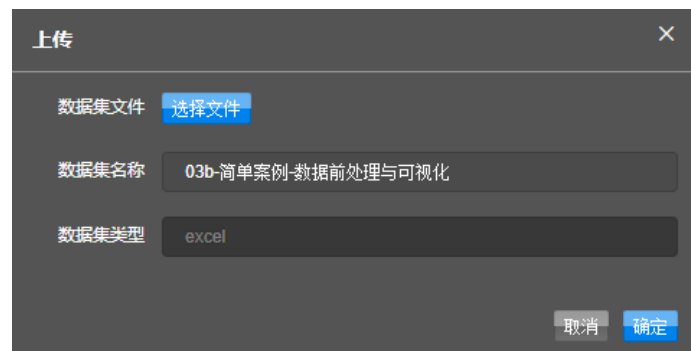


图355 上传案例原始数据

为方便显示，可以修改数据集的名称。具体操作方式为在左侧资源树的数据集上唤出右键菜单，点击其中的“重命名”选项，将两个数据集分别命名为“03a”与“03b”，如图 356 所示

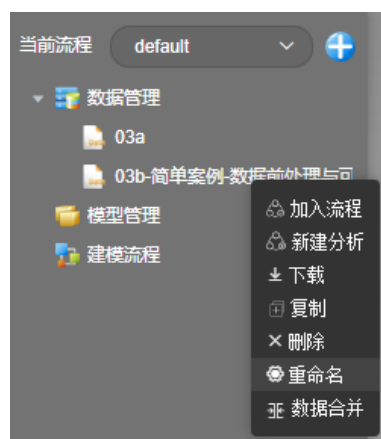


图356 重命名数据集

6.3.2.2 文件拼接

文件拼接功能可以将 2 个不同的数据集按照一定规则合并为 1 个数据集。具体操作方式如图 357 所示，在数据集 03a 上唤出右键菜单，选择其中的“数据合并-添加样本”选项。



图357 文件拼接-添加样本

在弹窗中，在“从数据管理列表”中选择数据集 03b，并点击继续，如图 358。



图358 选择添加数据集

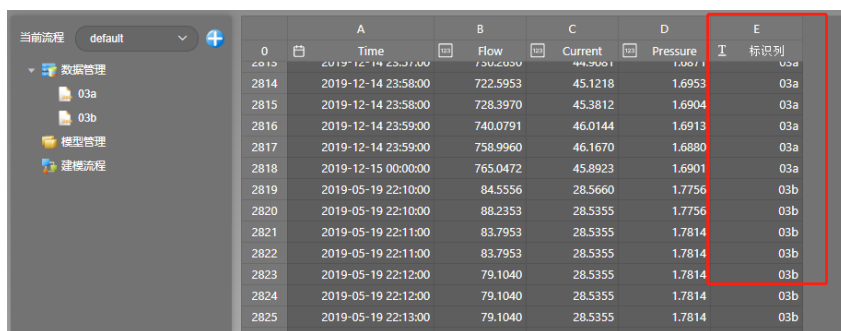
在新的弹窗中，可以看到右侧“合并后变量”中的变量名前都带有“(*)”标记，表明两个数据集的同名变量完成了自动匹配。勾选下方“追加列以标识数据源项”，并在文本框中填入“标识列”。最后点击确定，如图 359 所示。





图359 添加样本设定

点击刷新按钮以查看合并后的数据集，如图 360 所示，可以看到样本行数已增加，且在尾部增加了“标识列”以区分当前样本行的来源。



0	A	B	C	D	E
Time	Flow	Current	Pressure	标识列	
2814	2019-12-14 23:58:00	722.5953	45.1218	1.6953	03a
2815	2019-12-14 23:58:00	728.3970	45.3812	1.6904	03a
2816	2019-12-14 23:59:00	740.0791	46.0144	1.6913	03a
2817	2019-12-14 23:59:00	758.9960	46.1670	1.6880	03a
2818	2019-12-15 00:00:00	765.0472	45.8923	1.6901	03a
2819	2019-05-19 22:10:00	84.5556	28.5660	1.7756	03b
2820	2019-05-19 22:10:00	88.2353	28.5355	1.7756	03b
2821	2019-05-19 22:11:00	83.7953	28.5355	1.7814	03b
2822	2019-05-19 22:11:00	83.7953	28.5355	1.7814	03b
2823	2019-05-19 22:12:00	79.1040	28.5355	1.7814	03b
2824	2019-05-19 22:12:00	79.1040	28.5355	1.7814	03b
2825	2019-05-19 22:13:00	79.1040	28.5355	1.7814	03b

图360 合并后数据集

6.3.2.3 加载数据集

数据管理中的数据集默认为只读状态，仅支持复制等有限的功能。为开启完整的数据前处理功能，需提前加载数据集至本地内存。具体操作方式如图 361 所示。在顶部 Ribbon 菜单栏的数据处理 Tab 页下，点击“加载全部”按钮。完成后会提示“加载全部数据已完成”。



图361 加载全部数据

6.3.2.4 修改变量格式

导入的日期列与数值列可以被自动识别为正确的变量格式。在文件拼接操作中新生成的“标识列”的默认数据格式为纯文本格式。通过变量格式操作可以替



换为用户指定的格式，其操作方式如图 362 所示。在表头位置选中“标识列”（以蓝色高亮显示），唤出右键菜单并点击“变量格式”按钮，即可打开右侧弹窗。

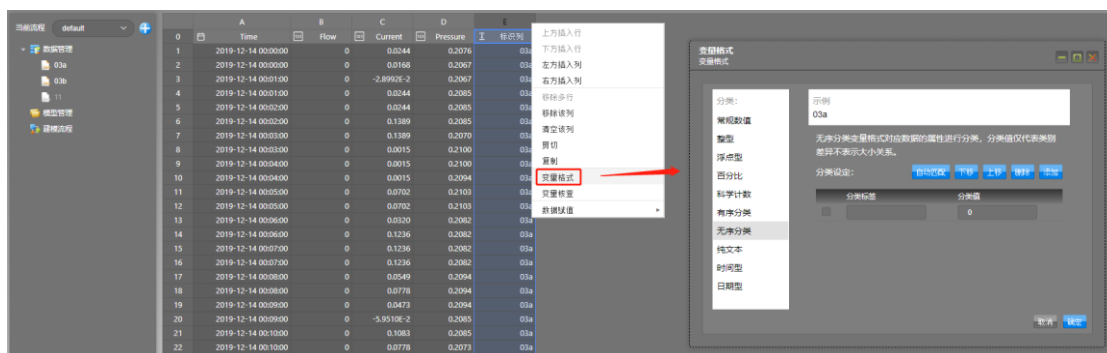


图362 打开变量格式弹窗

在变量格式弹窗中，选中左侧“无序分类”以标记该列为无序分类型变量。无序分类变量格式对应数据的属性进行分类，分类值仅代表类别差异不表示大小关系。点击右侧“自动匹配”按钮，即可根据当前列下内容自动填充分类标签并分配分类值，最后点击确定，如图 363 所示。可以看到对应表头的位置 ICON 已更换。



图363 变量格式设定

6.3.2.5 剔除无关样本

数据集中流量（Flow）列中为 0 的数据代表设备未运行，是与建模无关的样本，因此需要事先完成剔除。具体操作为：在表头位置唤出右键菜单，选择其中的变量核查选项，如图 364 所示。



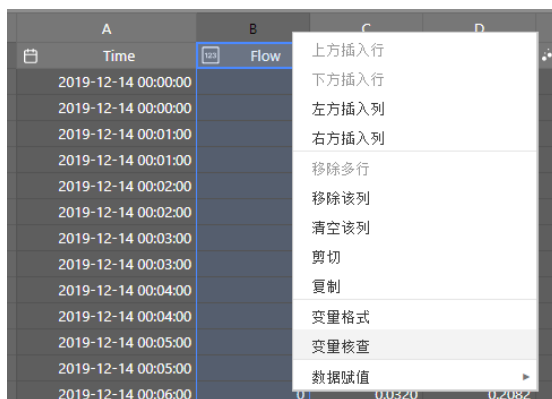


图364 变量核查

在弹出的对话框中，选择检查规则为“区间范围”，并设置方向和大小为“小于10”，之后点击确定，如图365所示。该步骤的含义为，筛选出当前列内所有小于10的数据。

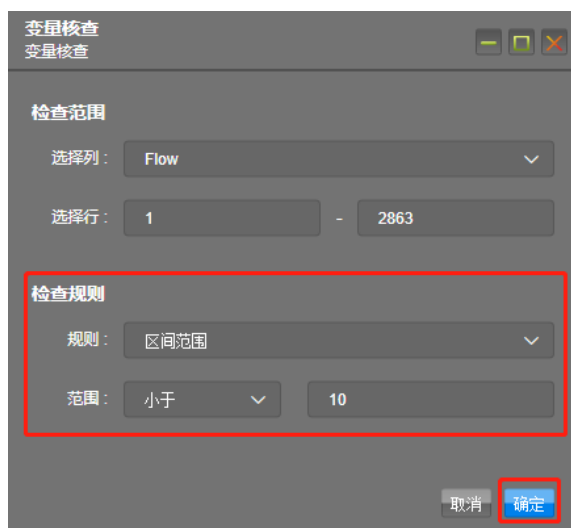


图365 设置检查规则

在下一个页面中，可以看到查找、删除和替换三个 Tab 页切换选项。进入“删除”页，选择第二个按钮“删除全部”，并点击确定，即可剔除所有无关数据。设置步骤如图366所示。





图366 删除无关样本

6.3.2.6 更新数据集

数据前处理中的操作需要点击“更新”按钮确认保存，之后才能继续完成如建模、可视化等操作。更新按钮在顶部 Ribbon 菜单栏，如图 367 所示。



图367 更新数据集

6.3.2.7 数据可视化

点击顶部 Ribbon 菜单栏中的“2D 散点图”按钮，X 轴变量与 Y 轴变量分别选择 Flow 与 Current，并点击确定，如图 368 所示。



图368 建立 2D 散点图

在新窗口中绘制了 2D 散点图，点击“y 轴变量”右侧的齿轮按钮，在“聚类分析”下拉选项中，选择标识变量为第 2 步中生成的标识列，可以看到散点图



中根据样本来源分为两种颜色展示。在 03b 对应的散点中，可以清晰地发现部分数据分布与 03a 中分布一致，体现为两种颜色的散点分布在同一条曲线上；而有另一部分 03b 对应的散点与其他散点距离较远，表明可能存在某种异常。分类标识的颜色可主动修改，如图 369 所示。

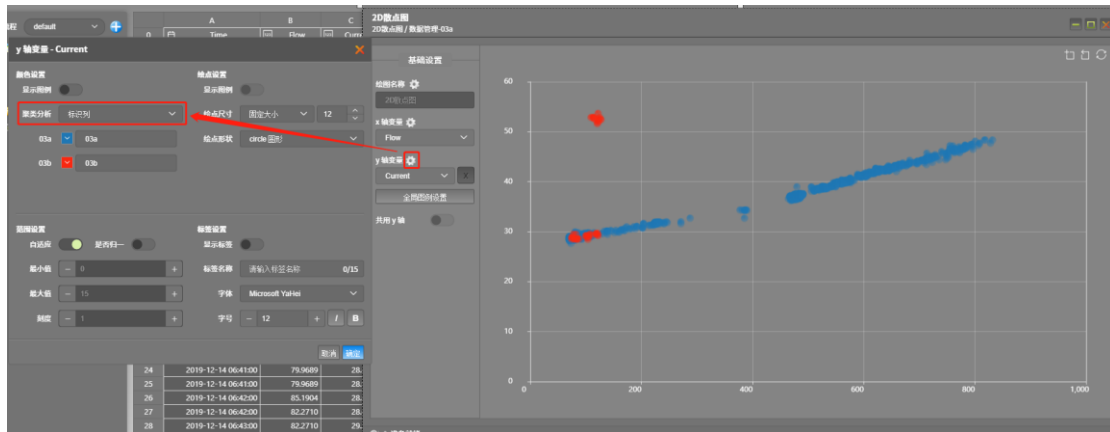


图369 数据可视化

至此，通过数据前处理和可视化的操作初步分析得到一定结论，后续可在建模流程中进一步详细分析。

6.4 专业案例 4——智能模型预警

6.4.1 案例描述

在本节中，将介绍 DTEmpower 在实际工程案例中的具体应用。

该案例来源于某电厂化学水处理系统反渗透过程，反渗透是采用膜法液体分离技术，利用膜两侧的压力差为动力，使溶液中的溶剂透过反渗透膜分离出来。本案例利用反渗透进水母管压力、一段入口流量、一段压差、一段产水流量、一段浓水流量等参数训练得出反渗透二段压差智能预警模型。其部分算例数据如图 370 所示。



	A	B	C	D	E	F	G
1	时间	反渗透进水母管温度	反渗透进水母管压力	反渗透入口电导率	反渗透入口ORP	反渗透一段入口流量	反渗透一段压差
2	2019/1/6 15:05	19.38194084	0.355872393	945.7463989	151.285141	77.15283966	101.4944763
3	2019/1/6 15:06	19.46063805	0.355172306	946.1856079	151.2744293	76.32886505	101.4655838
4	2019/1/6 15:07	19.45603561	0.354972333	946.7981567	151.2414398	76.25016785	101.3585205
5	2019/1/6 15:08	19.46993065	0.354972482	946.9187012	151.2236328	77.73858643	101.3567963
6	2019/1/6 15:09	19.45832443	0.354872376	947.3390503	151.2055664	77.9885788	101.3151016
7	2019/1/6 15:10	19.44211769	0.355572194	947.6312866	151.22258	76.90527344	101.4247208
8	2019/1/6 15:11	19.47691727	0.355272084	947.6959839	151.2236023	77.01177979	101.3169327
9	2019/1/6 15:12	19.44911957	0.355572164	947.9072876	151.2405243	76.77325439	101.3487473
10	2019/1/6 15:13	19.45831108	0.355472118	948.0521851	151.2610321	76.67824554	101.52668
11	2019/1/6 15:14	19.49310493	0.355172127	948.2802124	151.2743683	75.57424927	101.3376846
12	2019/1/6 15:15	19.51620483	0.355625242	948.6803589	151.2456665	76.23842621	101.4934921
13	2019/1/6 15:16	19.4838028	0.355525166	948.9108887	151.2808228	77.97224426	101.2354126
14	2019/1/6 15:17	19.48840904	0.354925066	949.1709595	151.2687378	78.59954834	101.4967117
15	2019/1/6 15:18	19.54171181	0.354225099	949.3907471	151.2874451	78.35668182	101.3025894
16	2019/1/6 15:19	19.50461197	0.354125112	949.5249634	151.2949982	77.41461182	101.367691
17	2019/1/6 15:20	19.54631233	0.355325103	949.7633667	151.3580322	76.43773651	101.3102036
18	2019/1/6 15:21	19.50232315	0.354425013	949.9047852	151.344635	77.31734467	101.4413071
19	2019/1/6 15:22	19.54633331	0.355024993	950.1634521	151.3643951	76.4471817	101.4756927
20	2019/1/6 15:23	19.53242493	0.35512501	950.2922363	151.3957672	78.09066772	101.3249664
21	2019/1/6 15:24	19.50692749	0.35512498	950.5449829	151.3790588	77.11377716	101.217453
22	2019/1/6 15:25	19.52552605	0.355124861	950.671875	151.3805695	77.8846817	101.2296524

图370 案例4部分原始数据

6.4.2 操作流程

6.4.2.1 添加数据集

在工程树状资源菜单“数据管理”项上单击鼠标右键，在右键快捷菜单中单击上传按钮，在弹出窗口中选择本地磁盘中的数据文件后单击确定即可完成数据集的添加，如图 371 所示。



图371 上传案例原始数据

数据集添加完成后即出现在左侧资源菜单中如图 372 所示。



图372 数据集上传结果

6.4.2.2 加入流程

在已添加的数据集上单击鼠标右键，在右键快捷菜单中选择加入流程菜单，原始数据的载入即以数据读取节点的形式加入至建模流程中如图 373 所示。

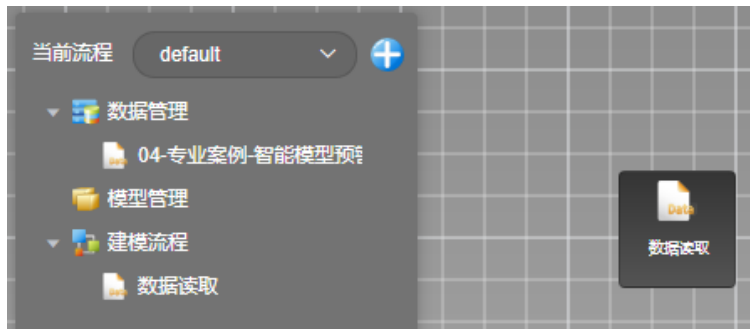


图373 加入流程

6.4.2.3 变量设定

将工具箱切换至数据操作，从中将变量设定节点拖拽至画布中，并连接数据读取节点，如图 374 所示。

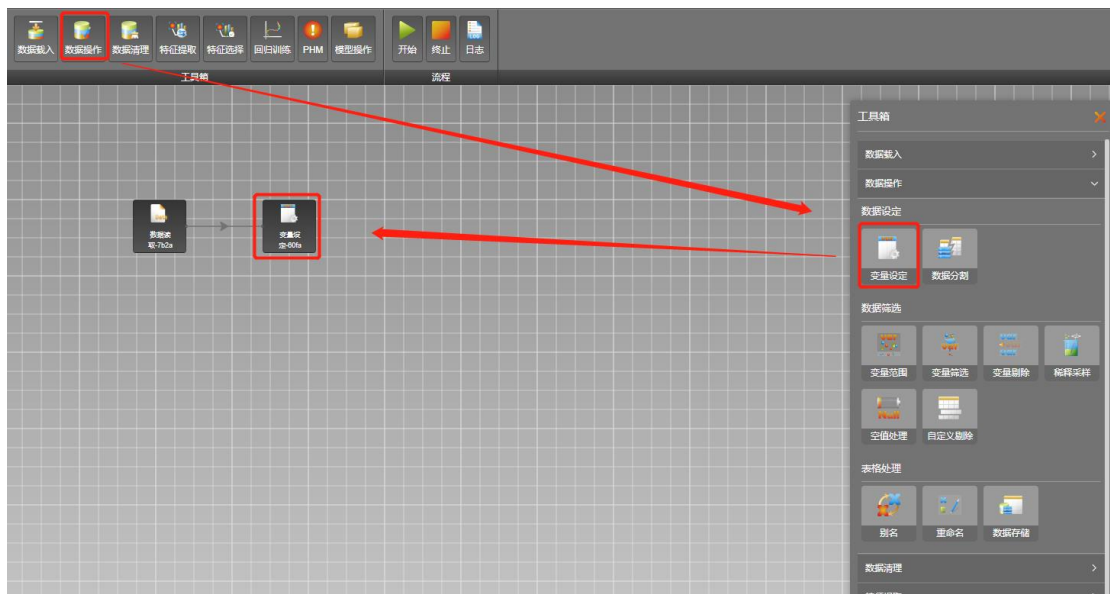


图374 添加变量设定节点

点击开始，当流程运行至变量设定节点时，在弹出配置窗口中将“反渗透二段压差”选为输出变量，其余（除时间）变量选为输入变量，如图 375 所示。





图375 变量设定配置

6.4.2.4 数据清理

将工具箱切换至数据清理，从中将 AIOD 节点拖拽至画布中，并连接变量设定节点，如图 376 所示。

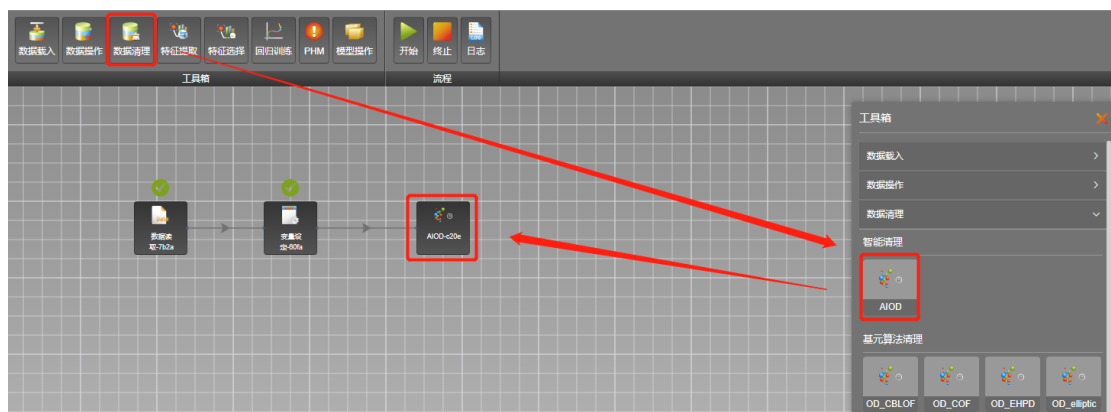


图376 添加 AIOD 节点

点击开始，在跳出配置弹窗中，左侧散点图绘制窗口的 Y 轴变量选择“_risk_scores”，X 轴变量选择“反渗透二段压差”，即可观察对应异常值的分布情况。在右侧表格中，设置风险阈值为 0.25 并点击筛选按钮，即可自动选中下方



表格中_risk_scores_大于 0.25 的样本。点击预览可观察异常样本的分布。最后点击右下角提交后完成异常点清理，如图 377 所示。

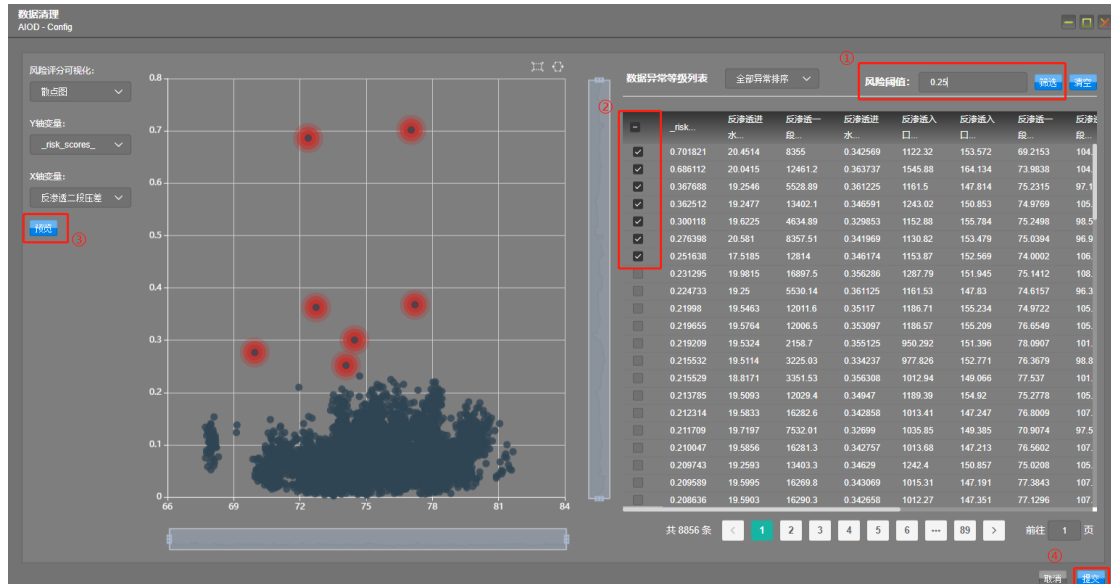


图377 AIOD 节点配置

6.4.2.5 敏感性分析

将工具箱切换至敏感性分析，从中将 GRA 节点拖拽至画布中，并连接 AIOD 节点，如图 378 所示。

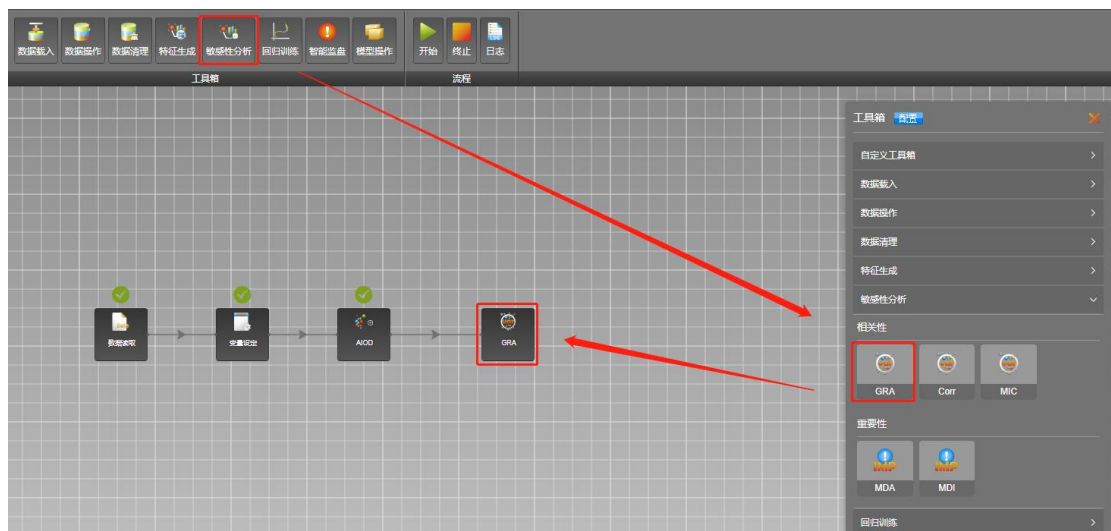


图378 添加 GRA 节点

点击开始，在跳出配置弹窗中，将反渗透二段压差的相关性阈值设为 0.7，并点击确定，如图 379 所示。在结果数据中，可以观察到已对特征变量进行了筛选。



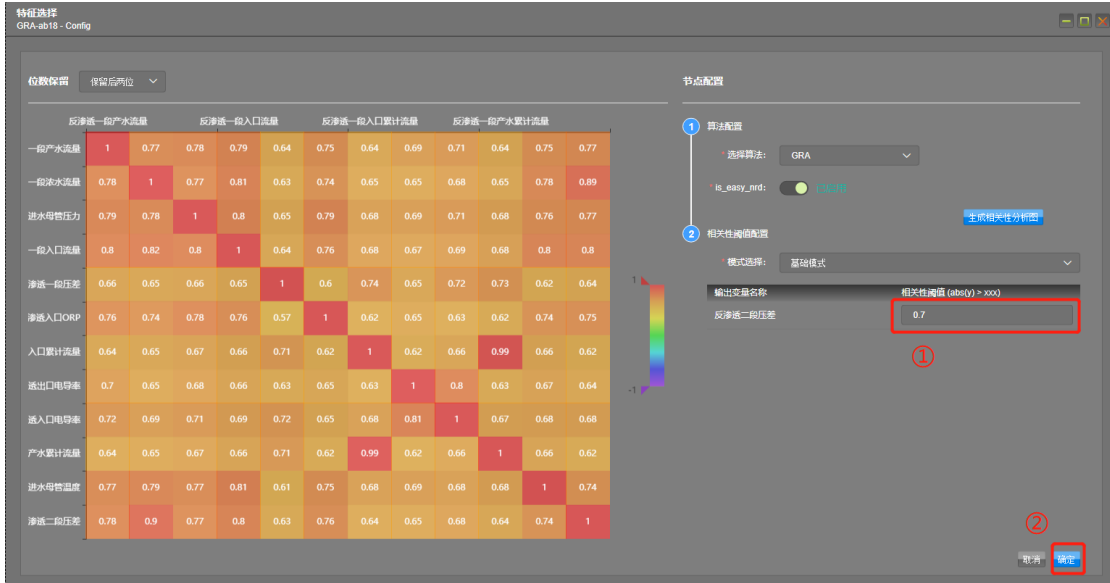


图379 GRA 节点配置

6.4.2.6 数据分割

将工具箱切换至数据操作，从中将数据分割节点拖拽至画布中，并连接 GRA 节点，如图 380 所示。

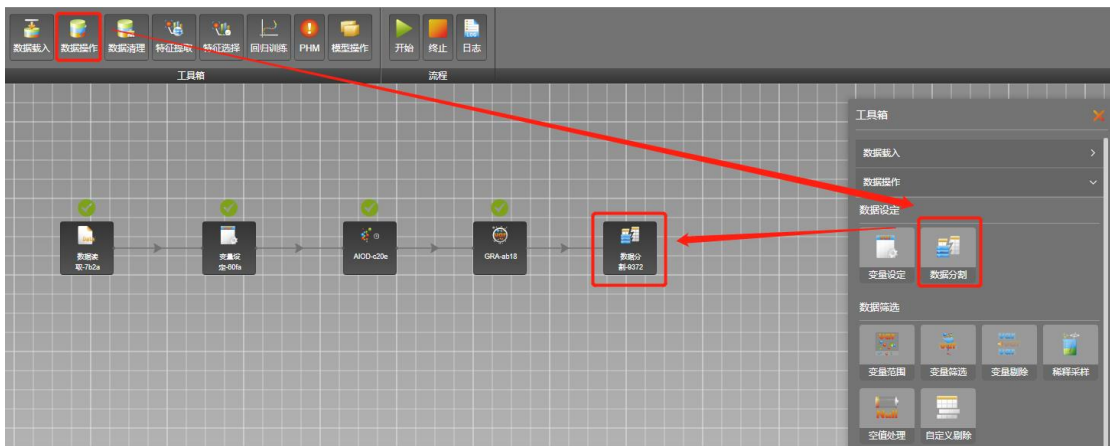


图380 添加数据分割节点

点击开始，数据分割节点将原始数据集随机分割为训练集与测试集，具体分割比例可在节点配置中调整，如图 381 所示。





图381 数据分割节点配置

6.4.2.7 模型训练

将工具箱切换至回归训练，从中将 BaggingDeFo 节点拖拽至画布中，并连接数据分割节点，如图 382 所示。

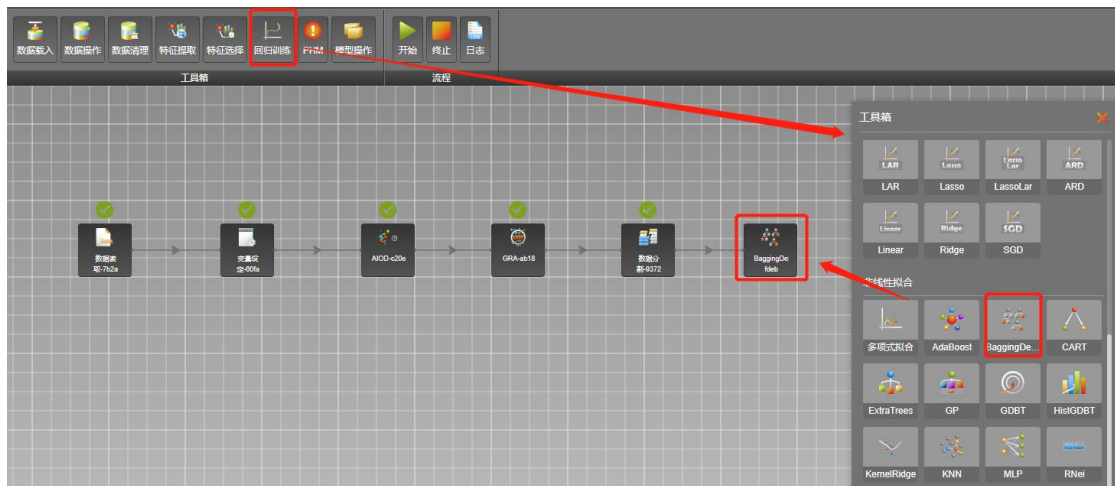


图382 添加模型训练节点

在 BaggingDeFo 节点配置中，开启交叉验证。点击开始，模型训练结束后，可在左侧资源树中查看模型详细信息。点击顶部菜单栏模型处理中的线性拟合或误差对比等，可观察模型训练效果，如图 383 所示。



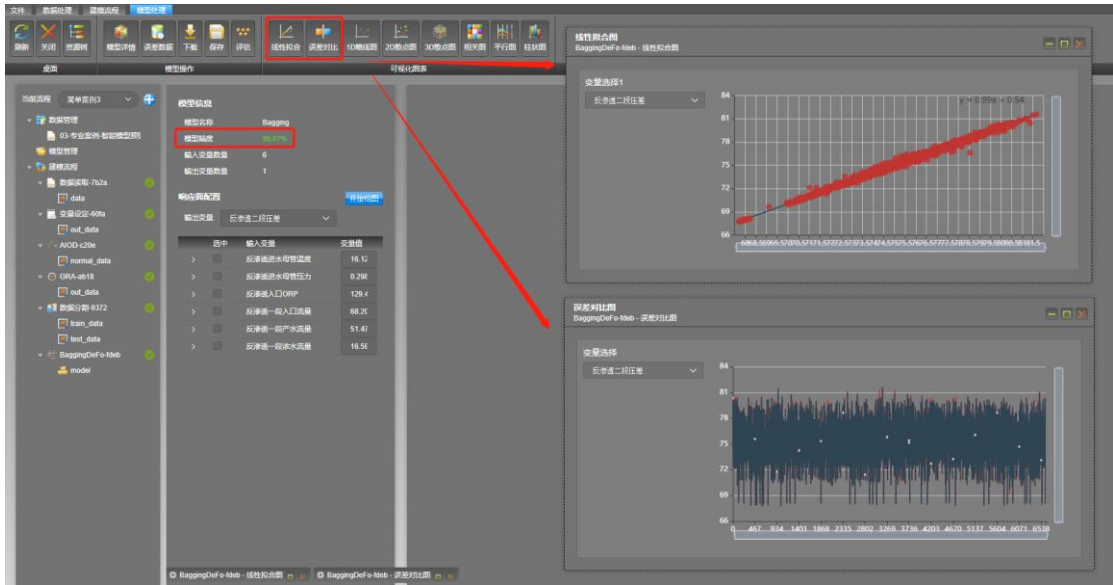


图383 查看模型训练效果

6.4.2.8 模型测试

将工具箱切换至模型操作，从中将模型对比节点拖拽至画布中，并同时连接数据分割与 BaggingDeForm 节点，如图 384 所示。

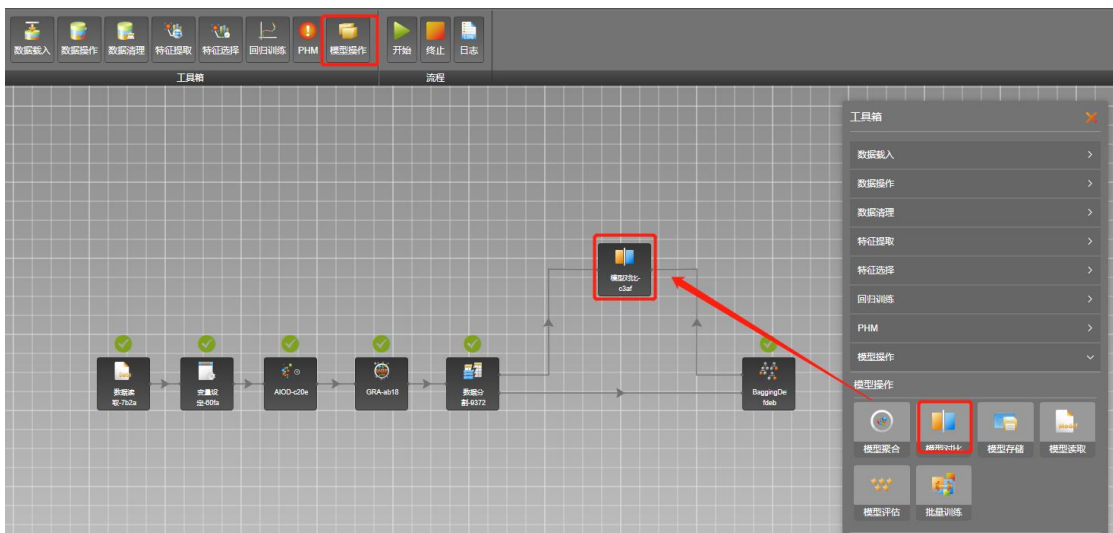


图384 添加模型对比节点

点击数据分割与模型对比之间的连线，在连线配置窗口中选择编辑，将输入值改为 test_data 并点击确定，如图 385 所示。



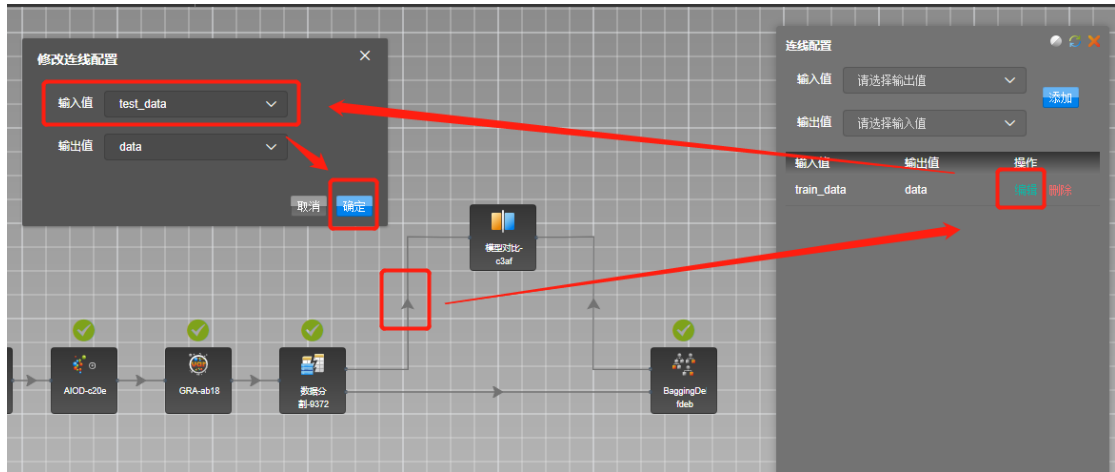


图385 模型对比连线配置

点击开始，运行结束后可查看模型测试结果，如图 386 所示。

名称	输出变量	r2	mse	mse	rmse	mse	mape	操作
Bagging	应渗透二段压差	0.9825112436403189	0.1943083151757794	0.08005266518052596	0.020688325203444043	0.28293579692312876	0.0025764352574895472	查看 清除

图386 查看模型测试结果

6.4.2.9 模型下载

当确认得到满意的模型训练效果后，可以通过在资源树的模型上点击鼠标右键下载至本地磁盘中，如图 387 所示。

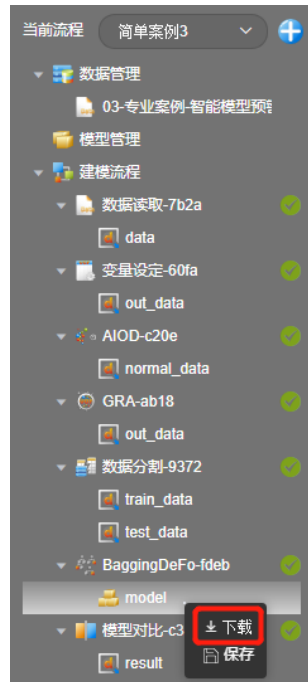


图387 下载模型



7 注意事项

7.1 浏览器兼容性问题

用户可以使用客户端或浏览器两种方式打开 DTEmpower。在浏览器使用模式下，用户务必使用 Chrome 浏览器接入系统。

7.2 表头要求

表头只能以字母或下划线开头；

表头中允许出现特殊符号（如~、!、@、和-等）；

表头中不允许出现空格；

DTEmpower 基于 Python 语言进行开发，因此表中的字符串不允许作为表头出现。

and	as	assert	break	class	continue
def	del	elif	else	except	finally
for	from	False	global	if	import
in	is	lambda	nonlocal	not	None
or	pass	raise	return	try	True
while	with	yield	np		

DTEmpower 数据库基于 SQL 设计，因此 SQL 关键字不允许作为表头出现，如 select, where, insert, update, delete, order, by, desc, asc 等（可参考 <http://www.postgres.cn/docs/9.3/sql-keywords-appendix.html>）。

此外，7.4 节中常用函数字符串不允许作为表头出现。

7.3 数据要求

对于参与模型训练的输入和输出变量，DTEmpower V2022R2 仅支持浮点数的形式。

若数据中包含时序变量，要求时序变量必须晚于 1970 年。



7.4 表达式函数列表

平台在变量表达式中支持如表所示的常用函数。

数学函数	表达式中的形式
常数	
π	@pi
e	@e
基础函数	
x^2	sqr
x^3	cube
10^x	10^
2^x	2^
$ x $	abs
\sqrt{x}	sqrt
$\sqrt[3]{x}$	cuberoot
$\log x$	lg
$\log_y x$	log
$\lfloor x \rfloor$	floor
$\lceil x \rceil$	ceil
$x!$	fact
$y\sqrt{x}$	yroot
$\ln x$	ln
e^x	x
三角函数	
$\sin x$	sin
$\cos x$	cos
$\tan x$	tan
$\sec x$	sec
$\csc x$	csc
$\cot x$	cot
$\sin^{-1} x$	arcsin
$\cos^{-1} x$	arccos
$\tan^{-1} x$	arctan
$\sec^{-1} x$	arcsec
$\csc^{-1} x$	arccsc
$\cot^{-1} x$	arccot
$\sinh x$	sinh
$\cosh x$	cosh
$\tanh x$	tanh
$\operatorname{sech} x$	sech
$\operatorname{csch} x$	csch



coth		coth
\sinh^{-1}		arcsinh
\cosh^{-1}		arccosh
\tanh^{-1}		arctanh
sech^{-1}		arcsech
csch^{-1}		arccsch
coth^{-1}		arccoth
统计函数		
最大值		maximum
最小值		minimum
平均数		smean
中位数		median
方差		var
标准差		std
求和		ssum
变异系数		cov
多元		
最大偏差		md
求和		fsum
平均数		fmean
逻辑		
<		<
<=		<=
=		==
≠		!=
>		>
>=		>=
与		and
或		or
非		not

