

数据挖掘-空值处理

概述

空值处理节点是将空值替换为均值、最大频数或者用户自定义的值等，实现空值的填充或者过滤。



- [概述](#)
- [输入/输出](#)
- [参数设置](#)
- [示例](#)

输入/输出

输入	一个输入端口，用于接收数据集。
输出	一个输出端口，用于输出空值处理的结果。

参数设置

设置空值处理的参数：

>> 参数

属性帮助

选择列 *必填 ②

选择列

空值处理

指定值

值 ②

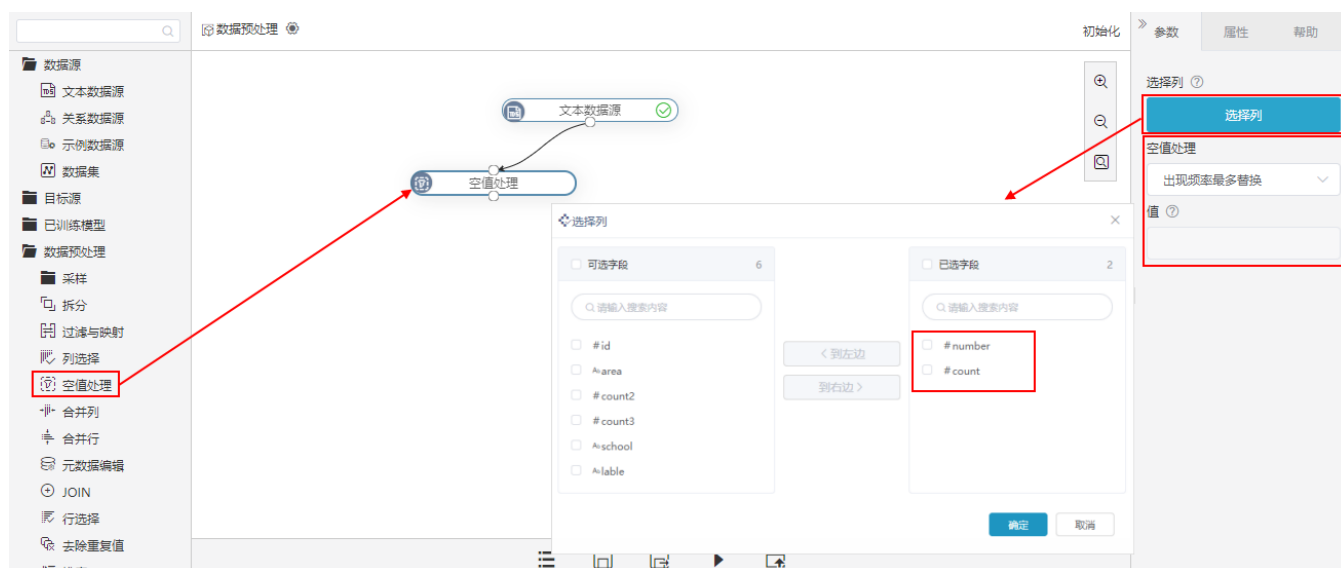
设置说明如下：

参数	说明
----	----

选择列	<p>用于选择进行空值过滤设置的列：</p> 
空值处理	<p>用于设置空值的处理类型：</p> <ul style="list-style-type: none"> • 最大值：使用本列字段的最大值替换空值，适用于数值型字段。 • 最小值：使用本列字段的最小值替换空值，适用于数值型字段。 • 平均值：使用本列字段的平均值替换空值，适用于数值型字段。 • 中位数：使用本列字段的中位数替换空值，适用于数值型字段。 • 按空值百分比删除列：若空值超过设置的百分比，则删除整列；适用于字符、数值型字段。 • 出现频率最多替换：使用本列出现频率最多的值替换空值；适用于字符、数值型字段。 • 指定值：使用设定的值替换空值，适用于字符、数值型字段。 • 过滤整行：只要选定行有空值，就删除该行。
值	<p>只有空值处理方法为以下两种时才需要进行值的设置：</p> <ul style="list-style-type: none"> • 按空值百分比删除列：输入百分比；表示空值单元格占该列所有单元格达到指定百分比时，删除该列。 • 指定值：输入具体指；表示以该具体值来显示空值。

示例

1、原先数据源存在空值，对其进行空值处理，选择列：number、count，空值处理为“出现频率最多替换”。



2、输出结果是将空值处理为该列出现频率最多的值，例如number出现频率最多的值为“4”，则空值经过处理后变成“4”；count出现频率最多的值为“100”，则空值经过处理后变成“100”。

id	area	number	count	count2	count3	school	lable
1	天河区	1	100	10	1	广州大学	!
2	null	2	null	20	1	中山大学	@
3	花都区	3	50	30	1	华南理工	#
4	null	3	null	40	1	华南农业	¥
5	番禺	4	30	50	1	广州外语	5
6	null	4	null	60	1	广州体校	sd
7	越秀	4	null	70	1	广州大学	wsa
8	荔湾	null	null	80	1	中山大学	null
9	null	6	null	90	1	华南理工	rfc
10	南沙	27	10	100	1	华南农业	jff

空值处理

id	area	number	count	count2	count3	school	lable
1	天河区	1	100	10	1	广州大学	!
2	null	2	100	20	1	中山大学	@
3	花都区	3	50	30	1	华南理工	#
4	null	3	100	40	1	华南农业	¥
5	番禺	4	30	50	1	广州外语	5
6	null	4	100	60	1	广州体校	sd
7	越秀	4	100	70	1	广州大学	wsa
8	荔湾	4	100	80	1	中山大学	null
9	null	6	100	90	1	华南理工	rfc
10	南沙	27	10	100	1	华南农业	jff