列选择节点

作者：张青松

目录

[1. 列选择 2](#_Toc433910773)

[2. 列选择操作示例 2](#_Toc433910774)

[2.1. 建立工作流 2](#_Toc433910775)

[2.2. 查看处理结果 5](#_Toc433910776)

# 列选择

在数据挖掘过程中，输入数据往往是来自数据库的数据表中，其中包含很多字段。然而仅仅只有一部分，或者是一小部分字段对我们感兴趣的数据挖掘的结果产生影响，其他字段都是无用的或者是噪声字段。为了方便选择出感兴趣的字段，使不重要的字段不会影响分析，需要使用“列选择”节点，“列选择”节点位于节点树中的“数据集”分组下的列处理下。

# 列选择操作示例

## 建立工作流

和其他数据处理节点一样，“列选择”节点需要一个有输出数据的节点来提供输入数据，本文中以“文件”节点为例，使用“文件”节点读取数据并连接到“列选择”的输入数据端口。工作流图如下图1，并且查看输入数据如图2。

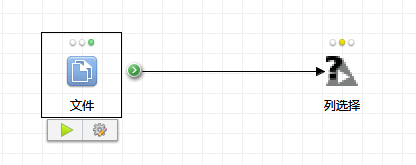


图1 列选择工作流

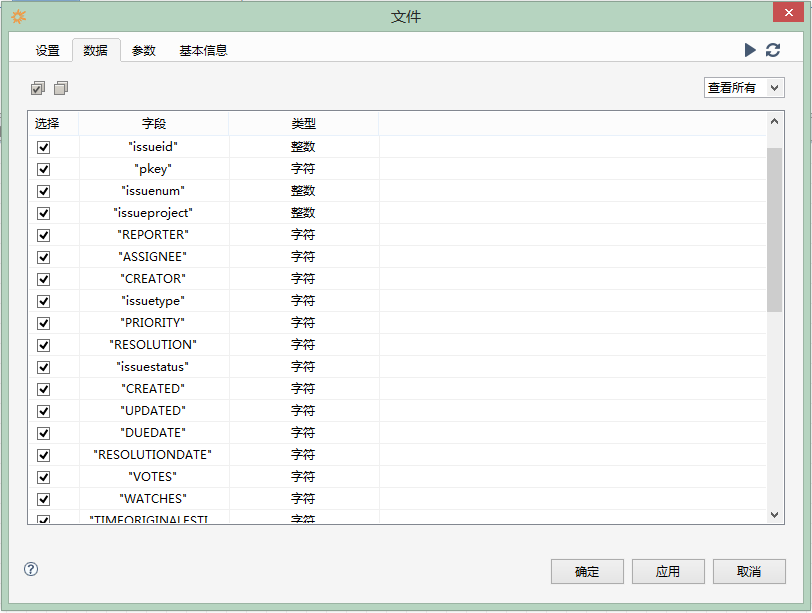


图2 输入文件字段

从文件节点的设置中可以看出，输入数据的字段数量较多，如果全部在某一个分析方法中，同时出现二三十个字段，不仅会影响分析者对分析算法的设置，也不利于分析者主观预测。

连接输入数据后，打开“列选择”节点的设置窗口，如下图3。

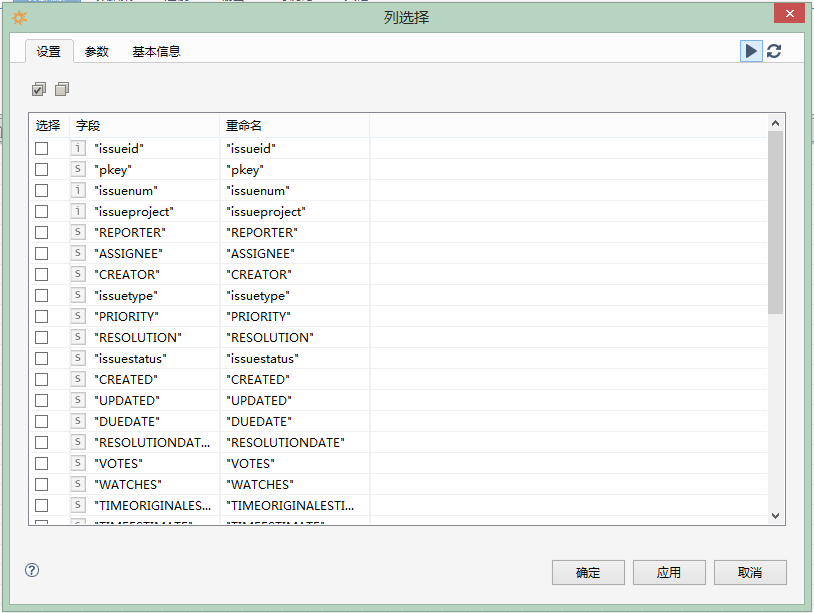


图3 列选择节点设置窗口

如图3所示，可以方便的从表中查看到各个输入字段的字段名，以及字段的数据类型。通过对应的复选框来选择该字段是否为后续挖掘过程感兴趣的字段。在表格中的第三列提供了对字段的重命名，对于不好辨认的字段名，手动输入自己容易识别的字段名，在输出数据中，替换原来的名称。如图4.

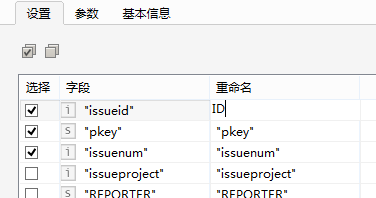


图4 选择字段并重命名字段

单击对应字段所在行的重命名单元格输入一个新的名称。注意，输入完成后必须在点击该窗口的其他任意位置，使输入名称的单元格失去焦点，触发重命名事件。

## 查看处理结果

如上节中的介绍完成设置，然后执行节点，并查看结果。对比输入数据图5和输出数据图6。

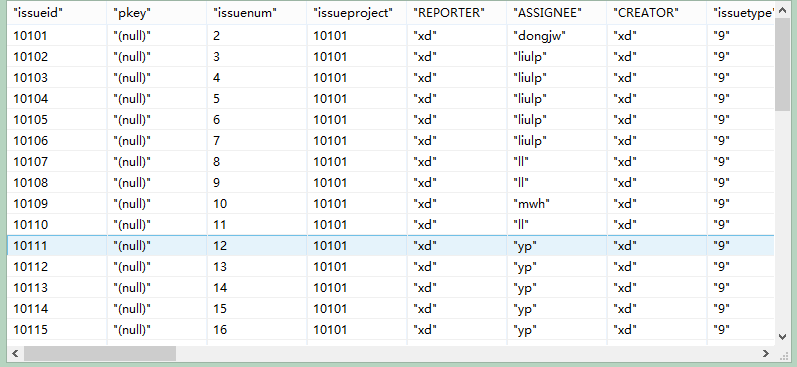


图5 列选择输入数据

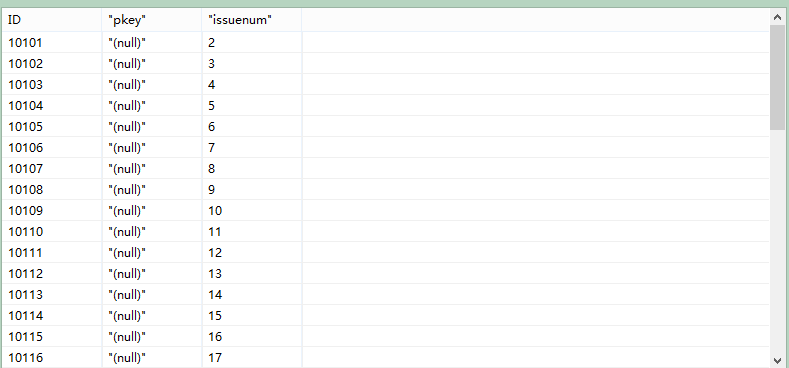


图6 列选择输入数据