3.4.2.8 记录去重

图标： 

描述：记录去重是去除数据表中的重复的行数据，只保留其中一行数据。

字段属性

特征列：必选。选择需要进行去重的列，勾选的字段可传入下个组件，如图 81 所示。

图 81

参数设置

无

输出

表结果：去重后的数据。

报告：无。

示例

- 勾选需要去重的数据，勾选的列将传入下个组件。如图 82 所示。
- 运行成功后，可右键查看数据，结果如图 83 所示。

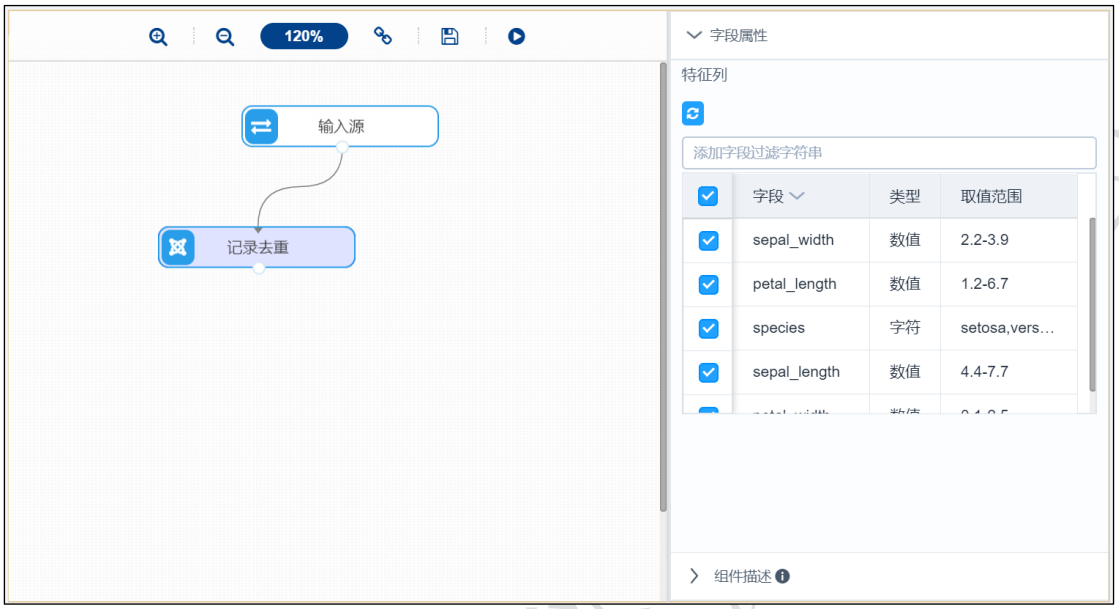

图 82

图 83

3.4.2.9 数据离散化

图标： 数据离散化

描述：某些模型算法，特别是某些分类算法如 ID3 决策树算法和 Apriori 算法等，要求数据是离散的，此时就需要将连续型特征（数值型）变换成离散型特征（类别型），即连续特征离散化。常用的离散化方法主要有三种：等宽法，等频法和通过聚类分析离散化（一维）。

字段属性

待离散化数据：必选。请选择数值型数据，如果勾选了非数值类型数据，则会自动过滤，