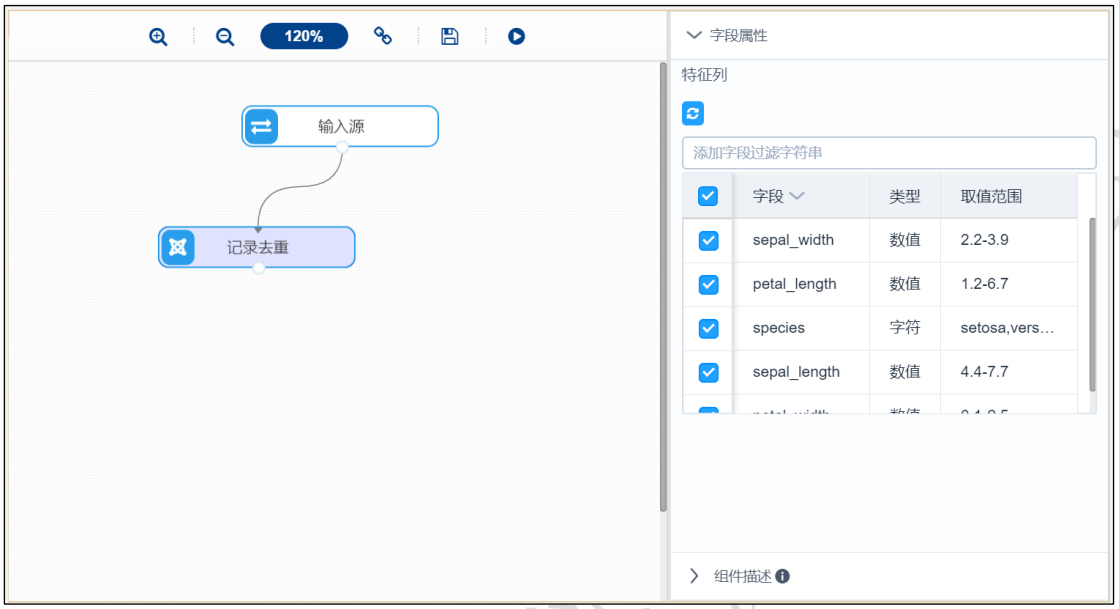

图 82

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，

下个组件可能无法获取所有列。勾选多列时，自动对每一列数据进行离散化如图 84 所示。

图 84

参数设置

离散化方式：选取要使用的离散方式，支持等宽、等频、聚类离散化，默认等宽。

离散个数：离散的个数，默认 2，如图 85 所示。

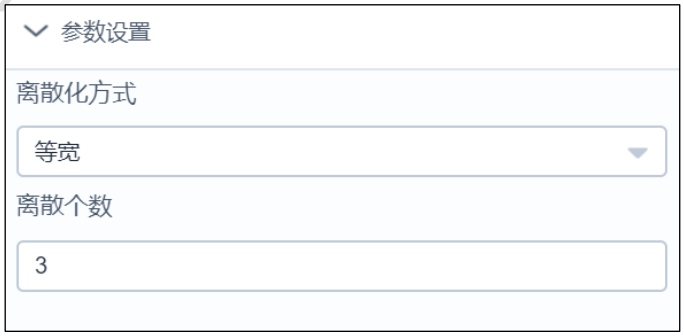

图 85

输出

表结果：对勾选的每一列进行离散化后的结果。

报告：无。

示例

下面对数据进行离散化。原数据如图 86 所示。

图 86

- 勾选需要进行离散化的数据，如图 87 所示。
- 选择离散化方式为等宽，离散个数为 2。如图 88 所示。
- 运行该组件，右击选择查看数据，如图 89 所示。

图 87

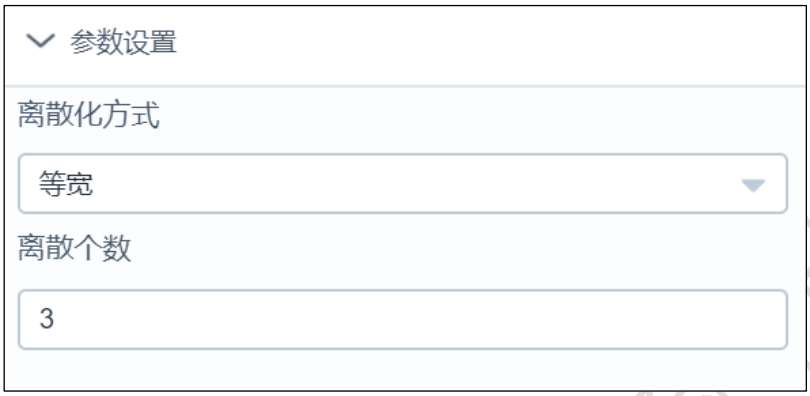

图 88

图 89

3.4.2.10 排序

图标： 

描述：根据某一列的顺序将所有数据重新排序。

字段属性

特征列：勾选的列必须包含关键字段。勾选的列将传入下一个组件。如图 90 所示。

关键字：必选，由该列值的顺序将待排序的数据重新排序。如图 91 所示。