通用算法-特征工程数据预处理：MinMaxScaler/StandardScaler项目开发说明：

MinMaxScaler：归一化方法

说明：是对原始数据的线性变换，使结果按列值映射到例如：[0 - 1]之间。转换函数如下

https://images0.cnblogs.com/blog/407700/201307/31105200-6bd5002661114e40ba1ee5d7d3377015.gif

min:列数据最小值

max:列数据最大值

StandardScaler：标准化方法

说明：这种方法给予原始数据的均值（mean）和标准差（standard deviation）进行数据的标准化。经过处理的数据符合标准正态分布，即均值为0，标准差为1，转化函数为

https://images0.cnblogs.com/blog/407700/201307/31105201-a6fe07c7a6764af0ac554988a3468917.gif

：列数据方差

：列数据均值

参数说明，在没有指定是MinMaxScaler/ StandardScaler特定参数的情况下，此参数在这两种数据处理方法中都适用。

|  |  |  |
| --- | --- | --- |
| 参数名称 | 参数使用说明 |  |
| action (train/predicitons) | train表示训练，predicitons表示预测 | 必传参数 |
| destPath | 结果数据集存放路径 | Predicitons过程，必传 |
| dataSetPathFile | 原始数据集加载路径 | 必传参数 |
| allCols | 原始数据集表头 | 原始数据集不含表头时，必传 |
| split | 原始数据集分隔方式，默认逗号分隔 | 当数据集是逗号分隔时，可以不传 |
| selectCols | 选择用于转化的列的列名，默认转化所有列 | 当想要转化所有列时，可以不传 |
| labelColName | 标签列的列名 | 当原始数据含有标签列时，需要传 |
| flag | 是否返回转化列的表头，默认不返回 | “ture”表示返回，false表示不返回 |
| modeFileName | 模型存放/加载路径 | 必传参数 |
| minValue | 设置转化数据最小值，默认0.0 | MinMaxScaler算法的个性化参数 |
| minValue | 设置转化数据最大值，默认1.0 | MinMaxScaler算法的个性化参数 |

例如：

a.调用训练接口：

{

"action": "train",

"dataSetPathFile": "/xx/xx/xx/ train.csv",

"labelColName": "label",

"modeFileName": "/xx/xx/xx/model",

"selectCols": "A,B,C,D,E ",

"split": ","

}

b.调用预测接口

{

"action": "predictions",

"dataSetPathFile": "/xx/xx/xx/ test.csv ",

"labelColName": " label ",

"flag": "true",

"modeFileName": "/xx/xx/xx/model ",

"destPath": "/xx/xx/xx/resultData",

"selectCols": "A,B,C,D,E ",

"split": ","

}