

# 最新！数据处理除了 SPSS, Origin, 还可以用这个

无论是实验处理数据，写文章，参加数学建模还是其他竞赛，我们在生活中经常需要处理数据，比如，目前文献、实验中常用的关系热力图、三元相图（图 1）、缺失值处理、线性回归等。

从力学性质三元相图我们可以得到不同元素比例与强度的关系，三元相图对于合金的制备有非常重要的指导作用。如图 1 所示 Cr 浓度是强度的主导因素，Co 浓度增加强度增加，Cr 元素浓度相对较大时，强度/(密度×价格)较大。

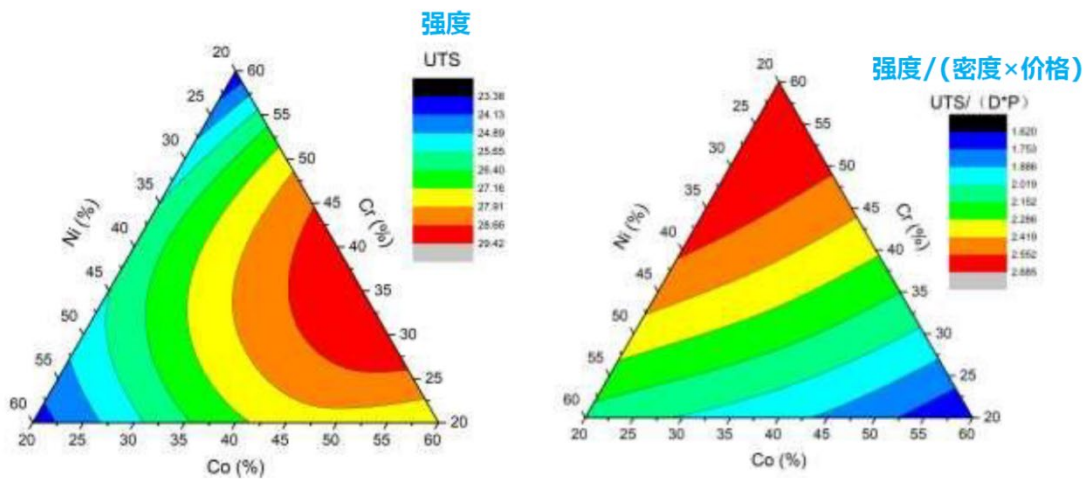


图 1 三元相图示例

这样的图是如何得到的呢？数据处理一般需要用到哪些软件？具体操作步骤是怎样的呢？

今天小编就带大家一起来看看吧！

下面我们以关系热力图为例，为大家介绍。

## 1. 关系热力图

如图 2 所示，该关系热力图用来表示元素之间的相关性分析。相关分析是研究两个或两个以上处于同等地位的随机变量间的相关关系的统计分析方法。相关性分析主要用于：（1）判断两个或多个变量之间的统计学关联；（2）如果存在关联，进一步分析关联强度和方向。

## 常用软件：SPSS、Python、Matlab...

对于非数学或统计专业的人并不需要 SPSS 这些全部的功能，也不熟练 Python 的代码输入。但在做数据分析的时候又需要用到时应该怎么办呢？

接下来就为大家介绍一种新的方式——MatCloud+，可以快速得到关系热力图。该方法的优点：无需下载软件，无需编程基础，在线即可处理数据，对于应急和短期需要使用数据处理的朋友十分友好。

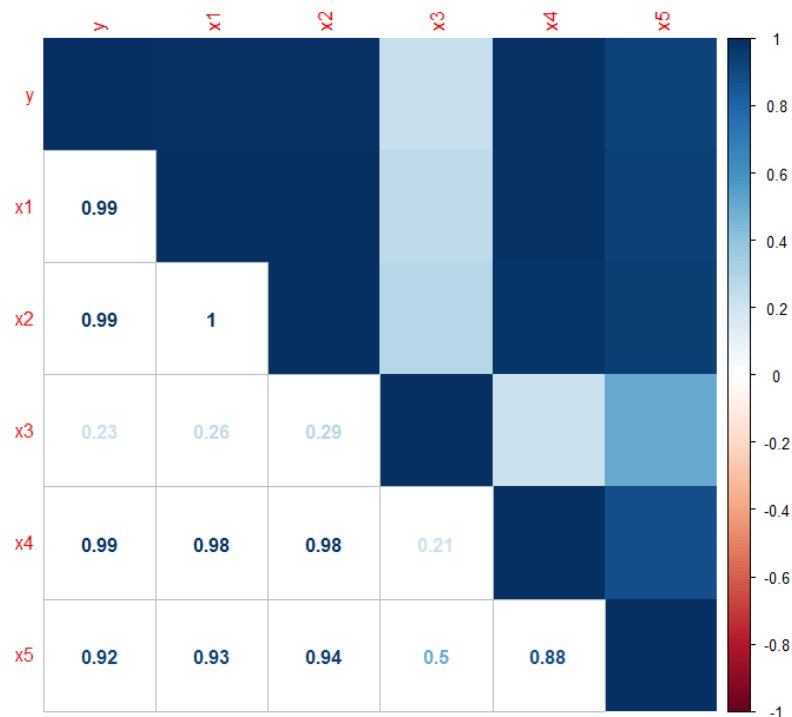


图 2 关系热力图示例

Matcloud+是怎样帮助大家完成关系热力图的绘制呢？首先，您只需要有一台能够上网的电脑即可。

- (1) 在百度搜索“MatCloud+”，进入平台的官网
- (2) 点进上方人工智能模块



图 3

(3) 点击新建工作流（也可先新建项目，在新建工作流）

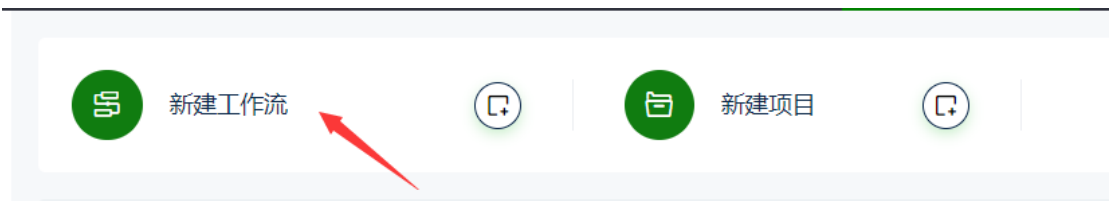


图 4

(4) 给您的工作流输入名称，并确认

### 新建工作流

名称 \*

描述

存储到 \*

图 5

进入新的页面之后就可以拖拽相应的组件来搭建 workflows 了，下面就是本次数据处理所需要的工作流及参数设置。

- a. 依次点击输入控制>数据输入，把【数据输入】拖到操作界面，点击【...】，上传需要处理的数据文档，保存，该界面支持查看，下载，删除。

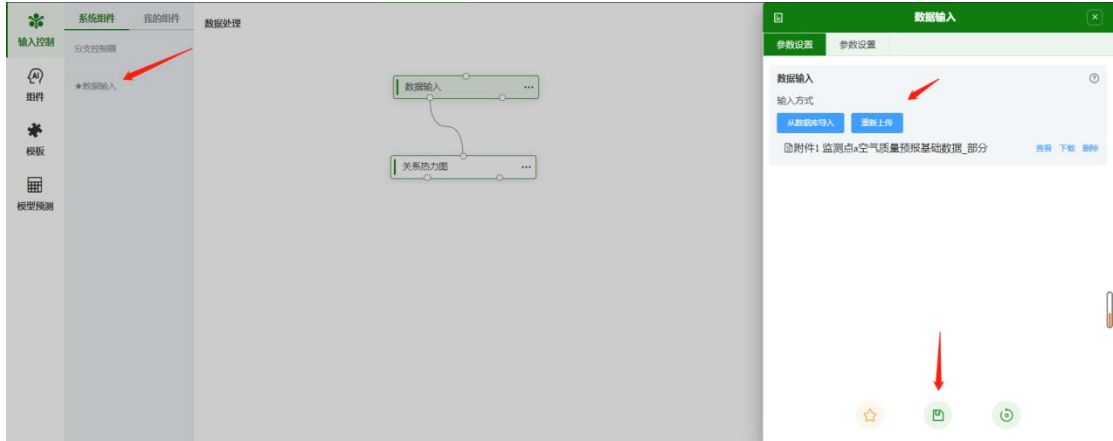


图 6

- b. 点击【组件】>【关系热力图】，拖拽至如图 7 所示，将【数据输入】与【关系热力图】两个组件进行连接。

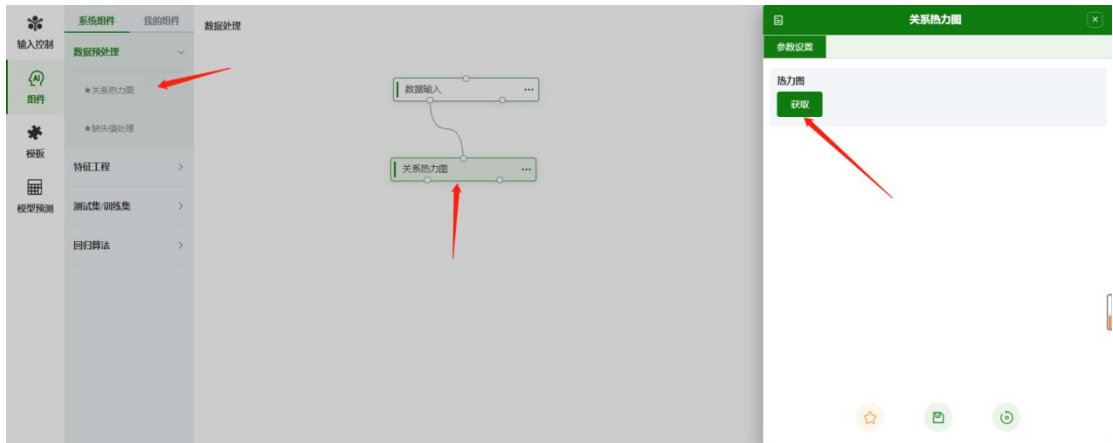


图 7

- c. 点击“获取”，即可得到结果，结果如下所示：



至此，通过**3步**便可得到相关性热力图。

MatCloud+将热力图处理过程中的复杂的步骤模板化，只需要拖拽数据预处理中的【关系热力图】即可，无需学习代码编写，不用安装软件，只要上传要处理的数据，几秒钟就能得到结果。无需去学习 Python 需要的代码，安装、下载 SPSS 软件，即可获得需要的结果，对初学者十分友好，上手较快。

## 2. 回归分析

在统计学中，回归分析 (regression analysis)指的是确定两种或两种以上变量间相互依赖的定量关系的一种统计分析方法。

回归算法是一种常用的数据分析方法，现在很火的人工智能最基本的算法就是回归算法。当你有很多材料制备的数据并想以这些数据为基础，来预测新材料的性能时就可以利用回归算法来得到相应的模型。同样的，MatCloud+支持快速获取这种模型。接下来为大家展示模型获取的步骤：

前几步和 1.关系热力图相同，无需下载软件、无需自己写代码，按照以下几个步骤即可完成。

- a. 依次点击输入控制>数据输入，把【数据输入】拖到操作界面，点击【...】，上传需要处理的数据文档。

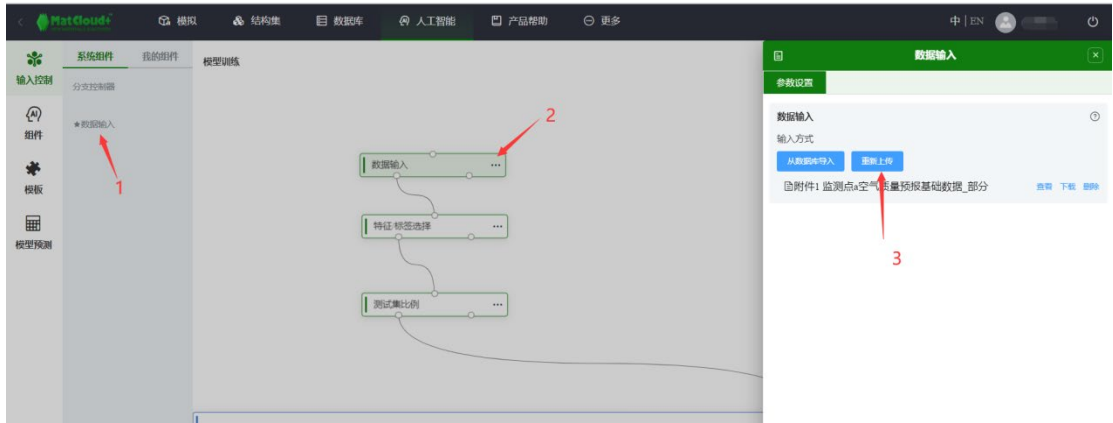


图 8

b. 点击：组件>特征工程>特征/标签选择，把【特征/标签选择】拖到【数据输入】下方，点击【...】，选择需要处理的数据标签，点击保存

c. 点击：组件>测试集/训练集>测试集比例，把【测试集比例】拖到【特征/标签选择】下方

并按如图 9 所示一一连接。

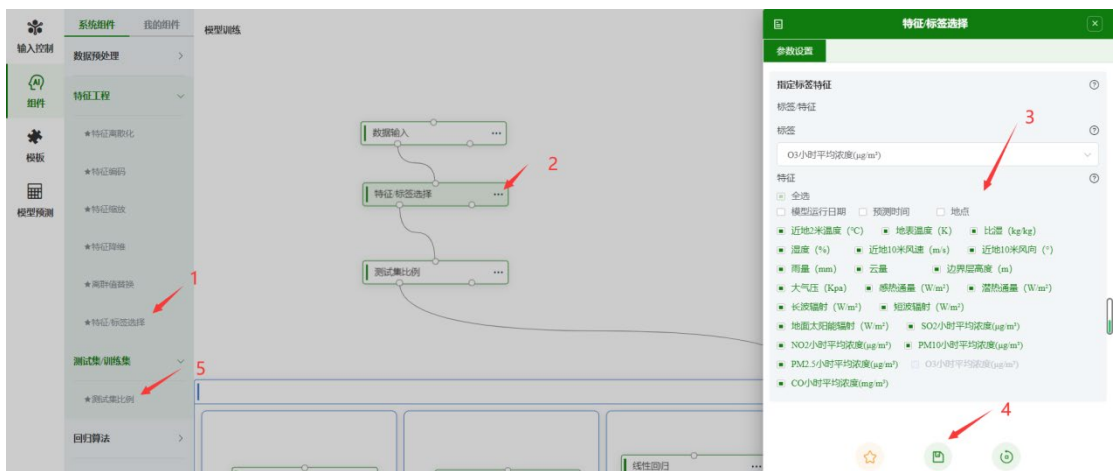


图 9

d. 点击输入控制>分支控制器，拖至【测试集比例】下方，根据需求添加分支的个数，不同分支可以同时进行不同的算法，我们可以在查看结果之后获来判断每个模型的好坏。

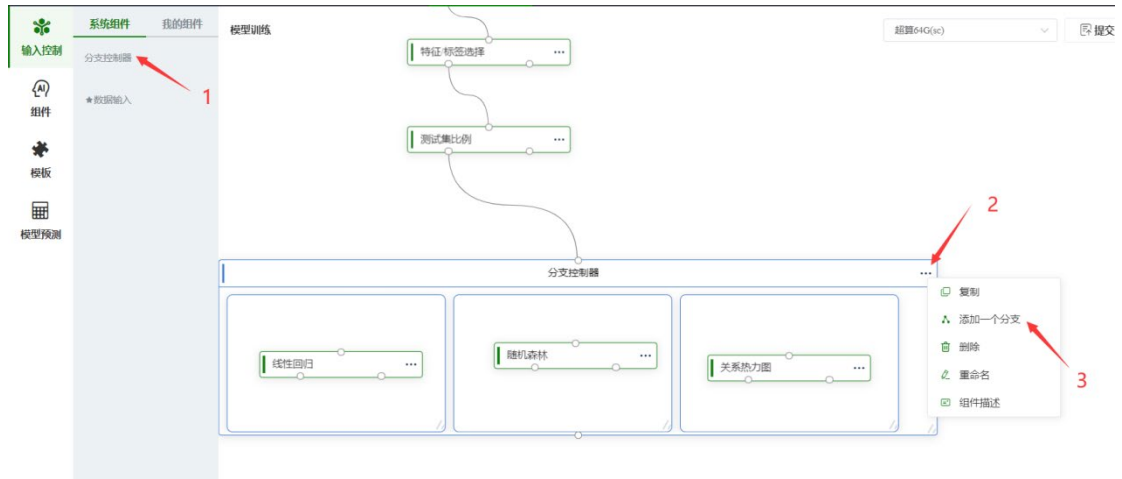


图 10

e. 点击组件>回归算法，有 K 近邻回归、XGBoost、决策树等...根据需求选择需要的数据处理方法，拖拽至【分支控制器】内即可

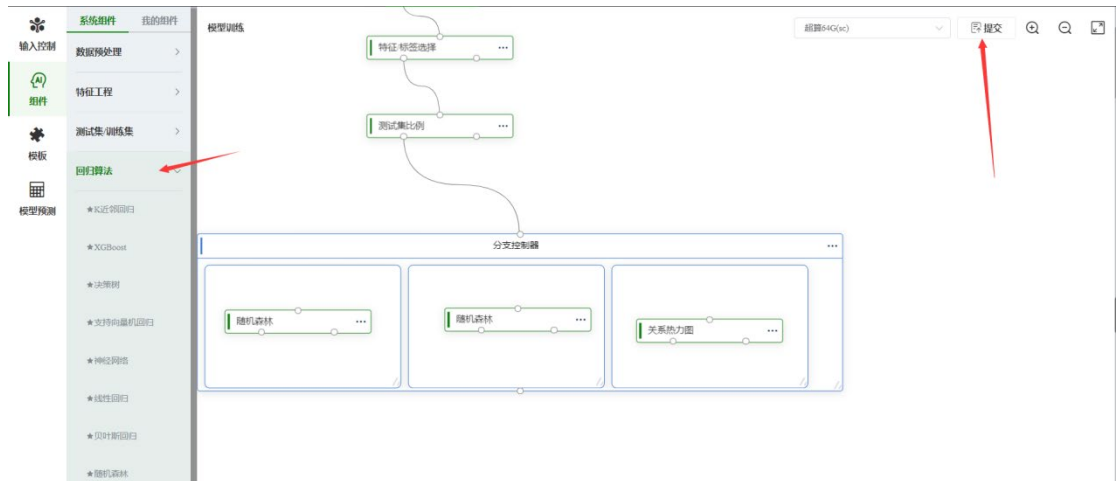


图 11

f.最后点击右上角的【提交】

等待运算完成。

运算结束后点击【...】，可查看结果。从图 12 中，我们可看出随机森林算法得到的结果，我们可根据  $R^2$  来判断预估模型的真实性和稳定性，如图 12 所示，现在  $R^2$  为 0.8722，则使用该模型之后残差的方差为原始值方差的 12.8%， $R$  方越接近 1，则预测值越接近于真值。并且在整个计算过程中，用到的模板均可以保存。

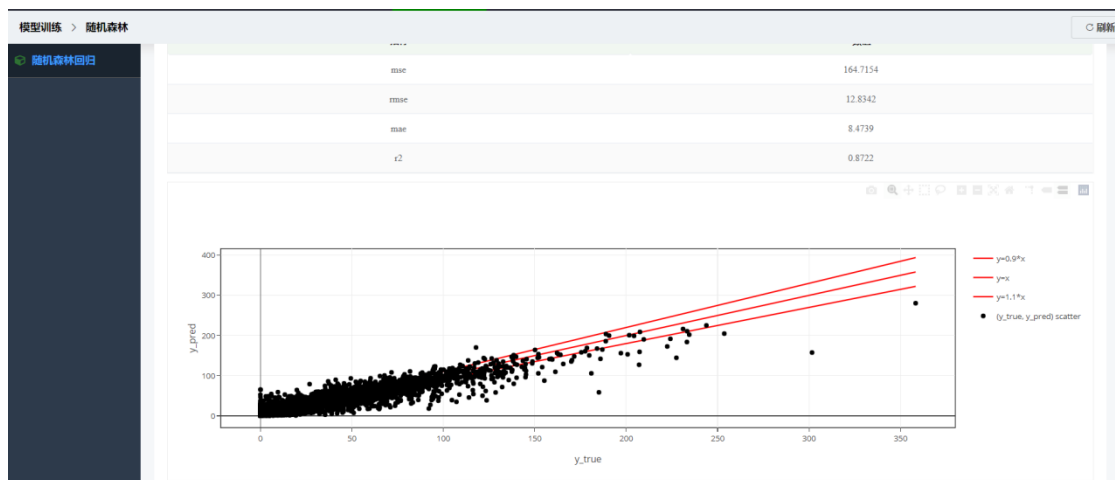


图 12

这里仅介绍了部分数据处理的使用方法，从整个过程中，相信朋友们已经体会到了该网站的优越性。事实上 MatCloud+平台还集成了大量的计算模块，如能带计算、弹性常数计算、光学性质的计算、数据处理等等。在不久的将来，MatCloud+平台有可能改变材料人的科研工作方式，从传统繁琐的计算中，将人解放出来，去更高效的完成其他工作。

之前平台有相关的培训课程，培训视频回放可在后台回复【matcloud+】获取“高通量多尺度材料计算和机器学习——高通量建模和第一性原理计算”视频及课件；回复【AI】可获取“LAMMPS+数据库+机器学习”视频及课件。

更多 Matcloud+教程可关注 **b 站迈高科技**。

更多动态请关注**迈高科技**微信公众号

