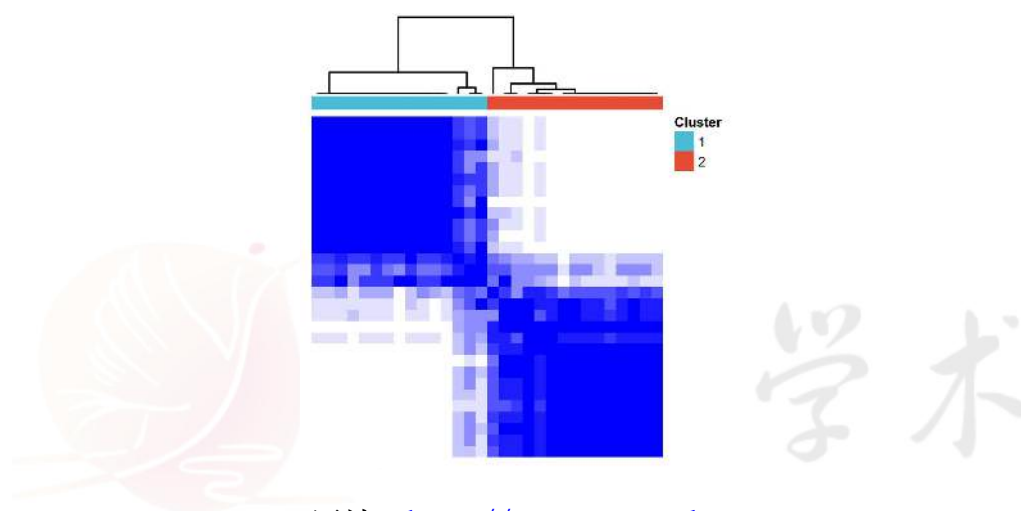


功能聚类 - 【一致性聚类】热图



网址: <https://www.xiantao.love>



更新时间: 2023.04.24

目录

基本概念	3
应用场景	4
主要结果	5
云端数据	6
参数说明	7
分析参数	7
颜色	8
风格	8
图片	9
结果说明	10
主要结果	10
补充结果	10
方法学	12
如何引用	13
常见问题	14

基本概念

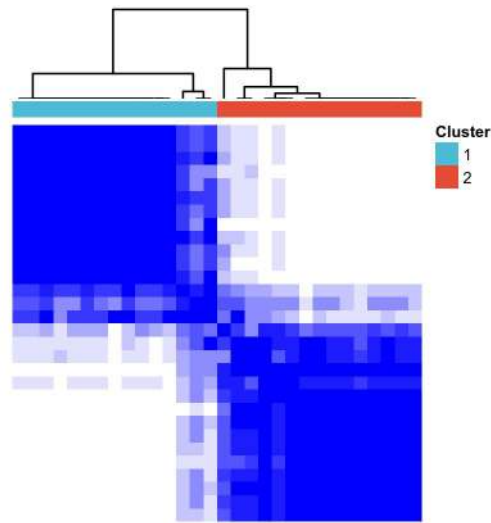
- 一致性聚类分析：使用不同的距离方法计算样本间的距离，对距离矩阵以重复抽样的方式来进行聚类分析，将多次聚类结果换算成概率，获得一致性聚类矩阵，从而验证聚类的合理性。
 - **聚类热图**：以热图的形式展示一致性聚类分析结果。
- hclust 层次聚类方法：
 - average, 两个类/群所有节点的平均距离作为两类的距离。
 - complete, 两个类/群中最长的节点作为两类的距离。
 - single, 两个类/群中最短的节点作为两类的距离。
 - centroid, 两个类/群中的中心点的距离作为两类的距离。
 - ward.D, 保证类合并时离均差平方和增量最下。
 - ward.D2, 在 ward.D 基础上进行了平方。

应用场景

一致性聚类是一种无监督聚类方法，可以将数据集中的样本区分成不同的亚型，随后对亚型进行比较分析。根据所选定的 K 值通过聚类热图的形式展示一致性聚类分析结果。



主要结果



一致性聚类热图：该图由三部分构成

- **主体**是由一致性分析矩阵数据绘制的热图，行和列均为样本，数值在 0 - 1 之间，颜色的深浅代表样本的聚类概率，白色代表样本不可能聚在一起。
- **热图上方**分别是聚类树和样本的分群信息，不同的颜色代表不同的分群，并由参数 K 值决定分类簇的多少。

云端数据

云端数据

	记录名称	来源模块	时间	补充说明
<input checked="" type="checkbox"/>		一致性聚类分析 @1.0	2023-04-24 10:01:32	数据记录可以在历史记录中找到

这里的云端数据与历史记录汇总 一致性聚类 分析模块的数据记录是保持一致的，可以在历史记录中找到相应的数据记录。

根据需要可视化的项目 选择好对应的云端数据记录。默认使用最近生成的分析记录。



参数说明

(说明：标注了颜色的为常用参数。)

分析参数

分析参数

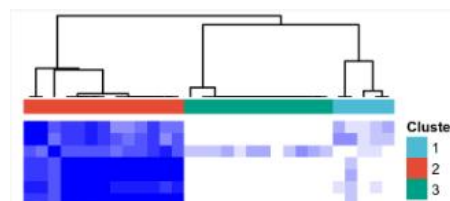
聚类方法

average

K值

2

- 聚类方法：此参数作为对一致性聚类得分矩阵进行层次聚类 hclust 的方法选项，默认使用 average。可以参考“基本概念”中层次聚类方法的说明。
- K 值：一致性聚类分析的分类簇数目，默认为 2，选项是 2-8 之间。一般选择 2-3 即可。



颜色



- 热图颜色：热图中的填充色选项，修改第一和第二色卡作为数值从小到大的渐变色，趋向于第一色卡（白色）即代表样本不可能聚在一起，反之趋向第二色卡代表样本为同一类的可能性越大。
- 聚类簇颜色：图中中间部分长方块的填充色选项，有多少个分类簇（K 值）会提取多少个颜色，最多支持修改 8 个颜色。受配色方案全局性修改。

风格



- 文字大小：控制整体文字大小，默认为 6pt

图片

图片

▼

宽度 (cm)

7

高度 (cm)

7

字体

Arial

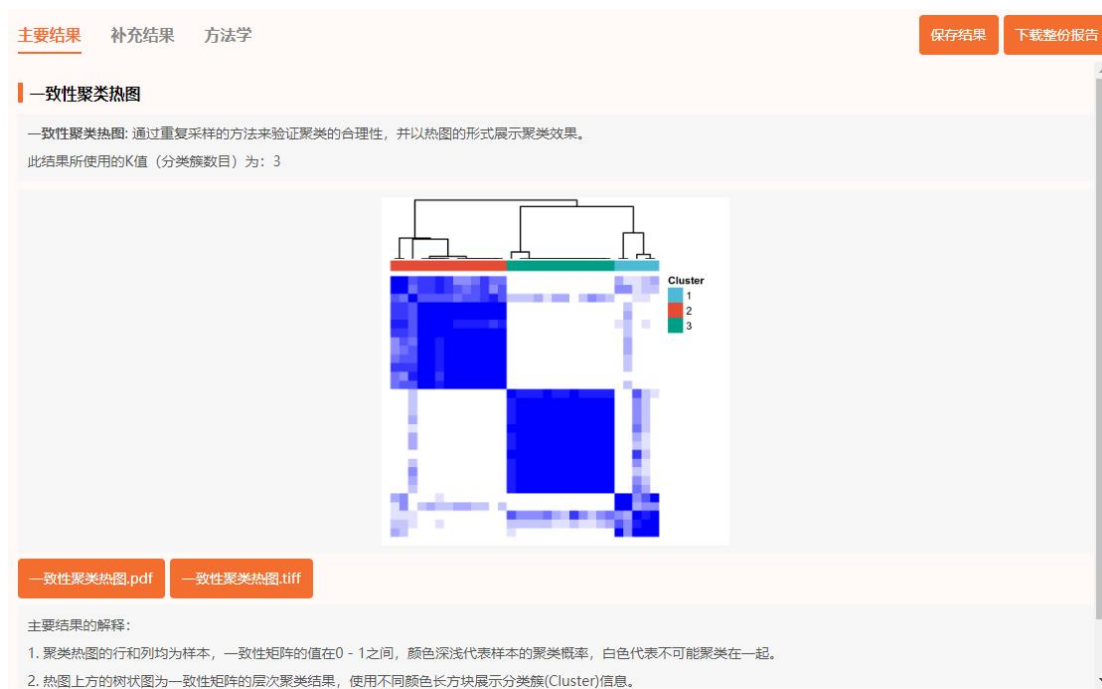
▼

- 宽度：图片横向长度，单位为 cm
- 高度：图片纵向长度，单位为 cm
- 字体：可以选择图片中文字的字体



结果说明

主要结果



主要结果格式为图片格式, 提供 PDF、TIFF 格式下载, 结果报告可以下载包括 pdf 以及说明文本的内容。

补充结果

分类簇情况

分类簇(Cluster)	样本数量
1	15
2	15

样本分群结果.xlsx

对给定K值的对称一致矩阵进行层次聚类, 获得每个样本对应的分类簇进行计数统计。
第一列为分类簇, 第二列为划分到该簇的样本数量。

此表格提供分类簇情况, 可以大致了解每个分类中有多少样本。

	A	B
1	Sample	Cluster
2	s1	1
3	s2	2
4	s3	1
5	s4	1
6	s5	1
7	s6	1
8	s7	1
9	s8	2
10	s9	1
11	s10	1
12	s11	1
13	s12	1

提供各样本的分群结果表格 [xlsx 下载](#)，一列为样本名，一列为分类簇。



方法学

所有分析和可视化均在 R 4.2.1 中进行

涉及的 R 包: pheatmap 包 (用于可视化)

处理过程:

- (1) 计算样本的距离矩阵, 通过重复多次抽样和聚类分析计算各样本在不同分类簇数目 K 下的一致性聚类分数
- (2) 根据选定的 K 值, 对一致性聚类得分结果进行层次聚类分析
- (3) 使用 pheatmap 包对得分和聚类结果进行可视化



如何引用

生信工具分析和可视化用的是 R 语言，可以直接写自己用 R 来进行分析和可视化即可，可以无需引用仙桃，如果想要引用仙桃，可以在致谢部分 (Acknowledge) 致谢仙桃学术 (www.xiantao love)。

方法学部分可以参考对应说明文本中的内容以及一些文献中的描述。



常见问题

1. 选择多少 K 值合适?

答:

一般 K 值选择 2-4 即可。

