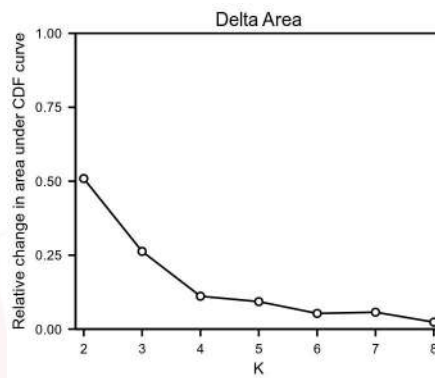


功能聚类 - 【一致性聚类】 Delta



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
点	7
线	8
标题	9
坐标轴	9
风格	10
图片	11
结果说明	12
主要结果	12
方法学	13
如何引用	14
常见问题	15

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。
- **Delta Area 图**：以点线图的形式展示各个 K 值 CDF 曲线下面积的相对变化。

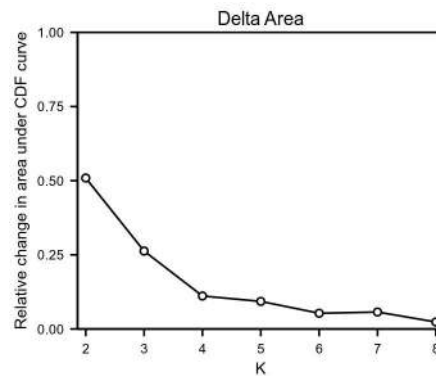


应用场景

通过一致性聚类分析可以获得不同 K 值下的一致性得分分布，使用 Delta Area 图可以直观评估和选择最合适的聚类数量 (K)。



主要结果



一致性聚类 Delta Area 图:

- 图中每一个点代表对应 K（分类簇数目）值下, K 减去 K-1 的 CDF 曲线下面积值, 横坐标是 K 值, 纵坐标是面积增长值。
- 当曲线或者点接近 0 时表明面积不再明显的增长, 此时可以考虑对应 K 值进行分类。



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

K值 8

- K 值：一致性聚类分析的分类簇数目，默认为 8，选项是 2-8 之间。一般使用默认即可，展示所有 K 之下的分布情况。

点

点

填充色

☐ ☐ ☐ ☐ ☐ ☐

描边颜色

☒ ☒ ☒ ☒ ☒ ☒

样式

圆形

大小

1

不透明度

1

- **填充色**：点的填充色颜色选项，有多少个分类簇会提取多少个颜色。支持修改 7 个颜色，分别代表 K 值 2-8。不受配色方案全局性修改。
- **描边颜色**：点的描边色颜色选项，默认黑色，有多少个分类簇会提取多少个颜色。支持修改 7 个颜色，分别代表 K 值 2-8。不受配色方案全局性修改。
- **样式**：点的样式类型，可选择 圆形、正方形、菱形、三角形、倒三角，默认为圆形。单选，选择后所有分组统一。
- **大小**：点的大小。
- **不透明度**：点的不透明度，默认为 1，代表完全不透明，此参数请输入 0-1 之间的值。

线



- **颜色**：连接点之间的连线颜色，默认黑色，支持修改 1 种颜色。
- **线条类型**：可以选择实线或者虚线。
- **线条粗细**：线的粗细。
- **不透明度**：线条的不透明度，默认为 1，代表完全不透明，此参数请输入 0-1 之间的值。

标题

标题

大标题

大标题内容

x轴标题

x轴标题内容

y轴标题

y轴标题内容

- 大标题：大标题文本
- x 轴标题：x 轴标题文本
- y 轴标题：y 轴标题文本
- 补充：在要换行的中间插入\n。如果需要上标，可以用两个英文输入法下的大括号括住，比如 {{2}}；如果需要下标，可以用两个英文输入法下的中括号括住，比如 [[2]]。

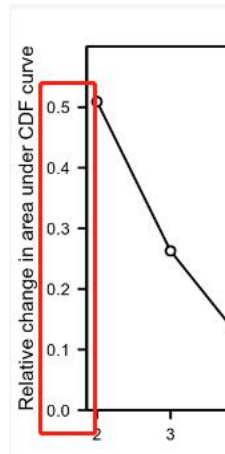
坐标轴

坐标轴

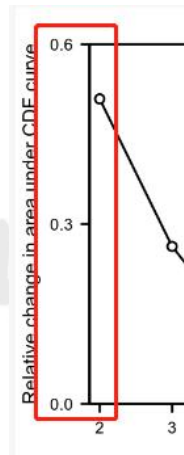
y轴范围+刻度

逗号隔开

- y 轴范围：可以控制 y 轴范围和刻度，提供 2 个值来控制范围。形如 -1, 1 (最小值和最大值不能超过 3, 调整过大可能会无作用)。
 - 如果只是想要修改范围，可以只输入两个范围值，比如 0,0.6



- 如果同时想要修改范围+刻度，可以输入比如：0,0,0.3,0.6,0.6。注意，此时最大和最小值会被当做范围值，不会作为刻度，如果需要刻度，需要类似于 0.6 那样同时写两次。



风格

风格
 ▼

边框
 ☒

网格
 ☐

文字大小
 6pt
 ▼

- 边框：是否添加主图外框
- 网格：是否添加网格
- 文字大小：控制整体文字大小，默认为 6pt

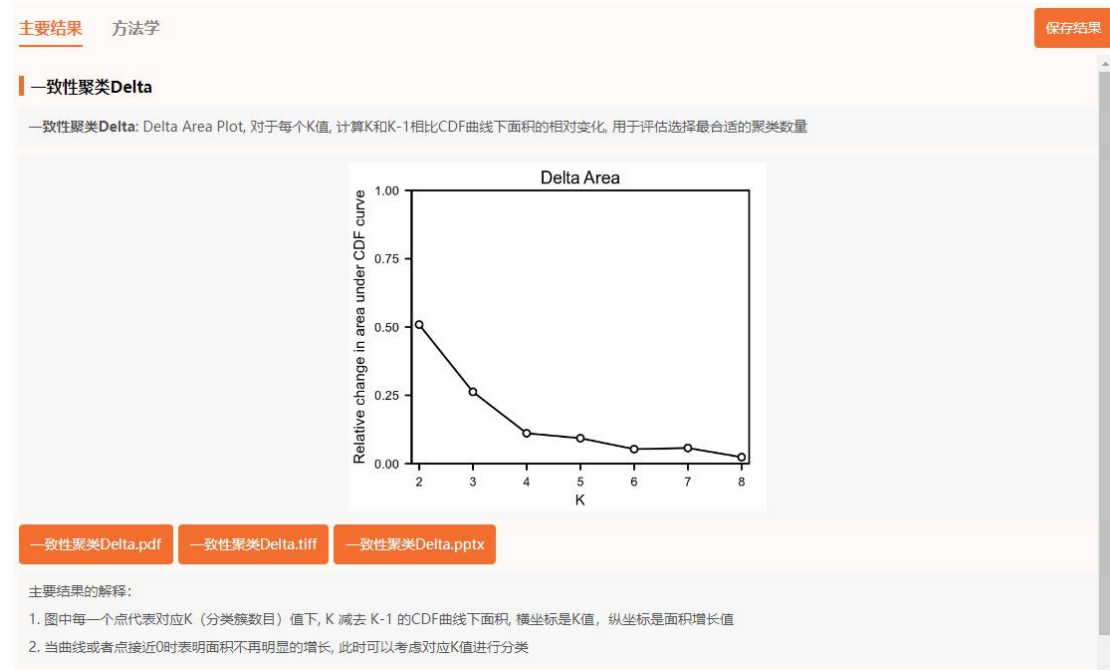
图片



- 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、PPTX 格式下载。

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: ggplot2 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) 使用直方图法计算不同 K 值的 CDF 曲面下面积, 取 K 减去 $K-1$ 的差值作为对应的 Delta 值
- (3) 使用 ggplot2 包绘制点线图进行可视化



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. 如何选择合适的 K 值?

答:

当曲线或者点接近 0 时表明面积不再明显的增长时，对应的 K 值较为合适。

