

分析报告

目录

任务 1 数据的预处理.....	2
1.1 产品通用名称规范化处理.....	2
1.2 计算总无机养分百分比.....	2
任务 2 肥料产品的数据分析.....	2
2.1 对复混肥料产品的分组.....	2
2.2 对有机肥料产品的分组.....	3
2.3 对复混肥料产品使用聚类算法分类.....	4
任务 3 肥料产品的多维度对比分析.....	6
3.1 对复混肥料产品登记数量的变化趋势的分析及可视化.....	6
3.2 分析广西和湖北两个省份组别的分布差异.....	6
3.3 计算企业之间的杰卡德相似系数矩阵.....	7
任务 4 肥料产品的多维度对比分析.....	7
4.1 提取出氮、磷、钾养分和有机质的百分比，以及肥料含氯的程度.....	7
4.2 提取各种原料的名称及其百分比.....	7

任务 1 数据的预处理

1.1 产品通用名称规范化处理

处理目标：

按照复混肥料（掺混肥料归入这一类）、有机-无机复混肥料、有机肥料和床土调酸剂 4 种类别对附件 1 进行规范化处理。

1.2 计算总无机养分百分比

处理目标：

计算附件 1 中各肥料产品的氮、磷、钾养分百分比之和，即总无机养分百分比。

任务 2 肥料产品的数据分析

2.1 对复混肥料产品的分组

分析目标：

从附件 2 中筛选出复混肥料的产品，然后将所有复混肥料按照总无机养分百分比的取值等距分为 10 组。并根据每个产品所在的分组，为其打上分组标签。

即可得到最终结果，部分数据如下图所示：

分析复混肥料产品的分布特点，绘制产品登记数量的直方图如下：

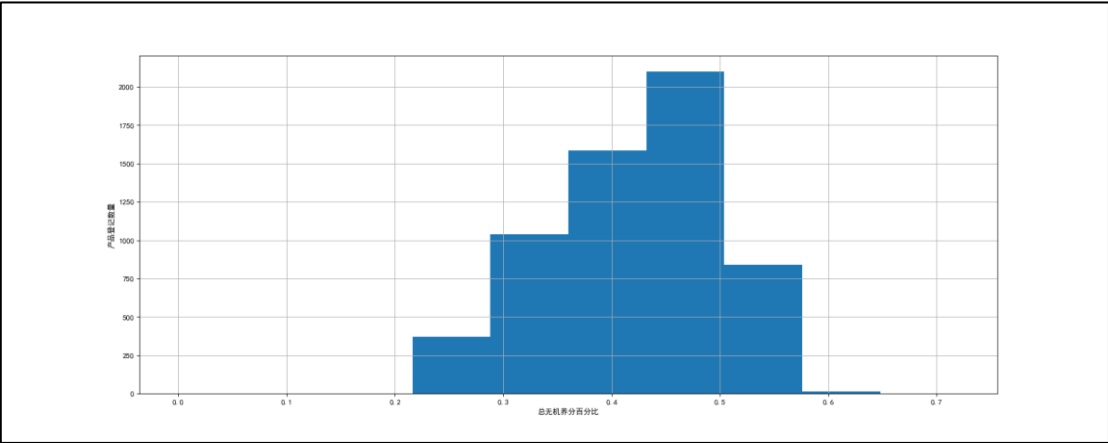


图 2.1.3 直方图

登记数量最大的前 3 个分组及相应的产品登记数量如下：

排名	一	二	三
分组标签	7	6	5
产品登记数量	2012	1501	1038

图 2.1.4 登记数量结果示意图

2.2 对有机肥料产品的分组

分析目标：

从附件 2 中筛选出有机肥料的产品，将产品按照总无机养分百分比和有机质百分比分别等距分为 10 组，并为每个产品打上分组标签。

绘制有机肥料产品的分布热力图如下：

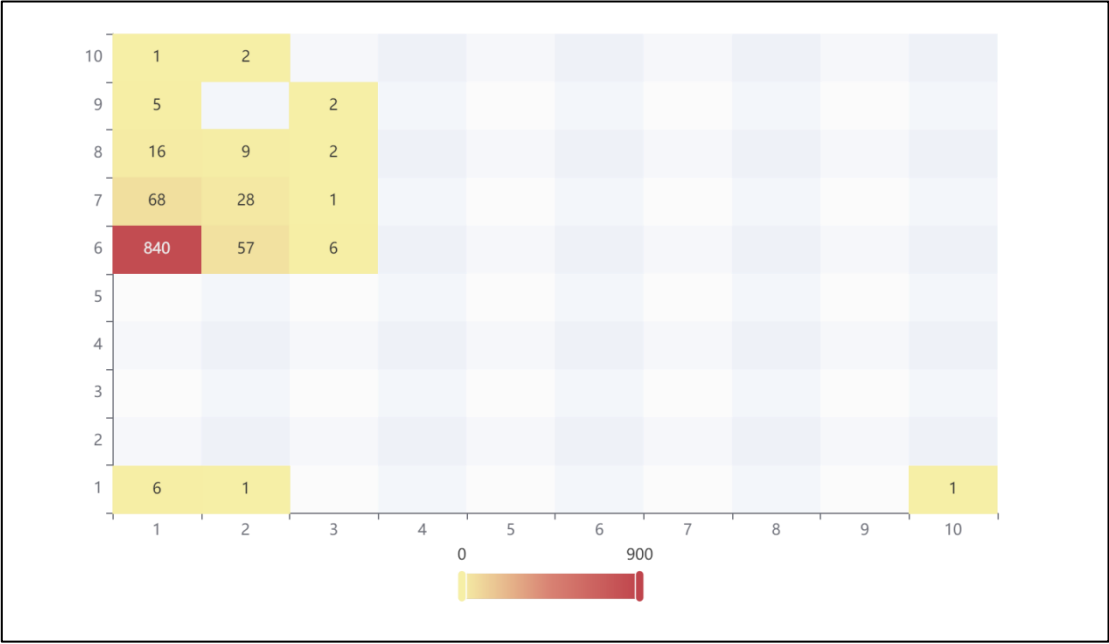


图 2. 2. 3 热力分布图

其中横轴代表总无机养分分组，纵轴代表有机质分组。

2.3 对复混肥料产品使用聚类算法分类

分析目标：

从附件 2 中筛选出复混肥料的产品，按照氮、磷、钾养分的百分比，使用聚类算法将这些产品分为 4 类。根据聚类结果为每个产品打上聚类标签。

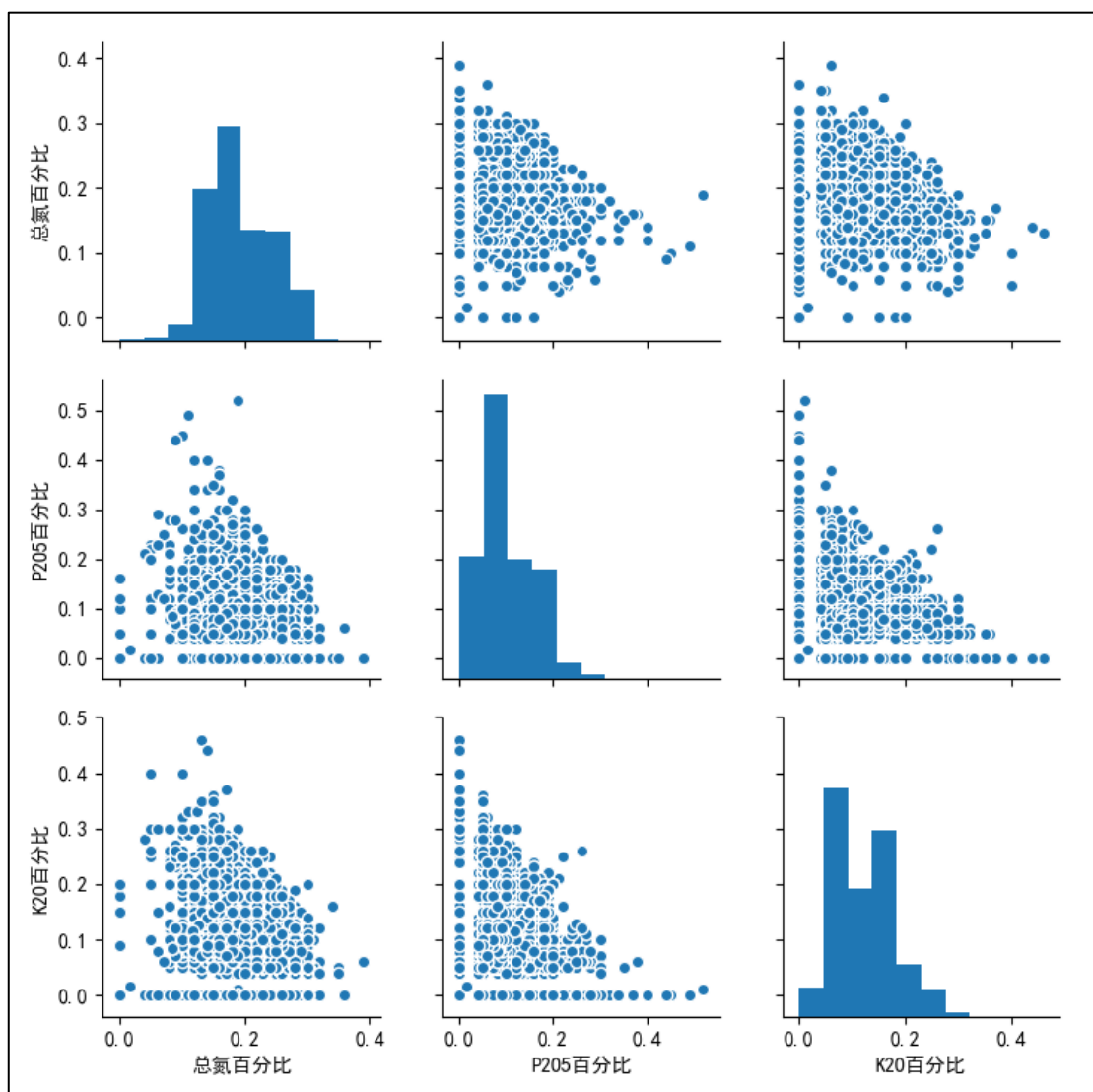


图 2.3.2 散点图矩阵

绘制聚类结果的雷达图如下：

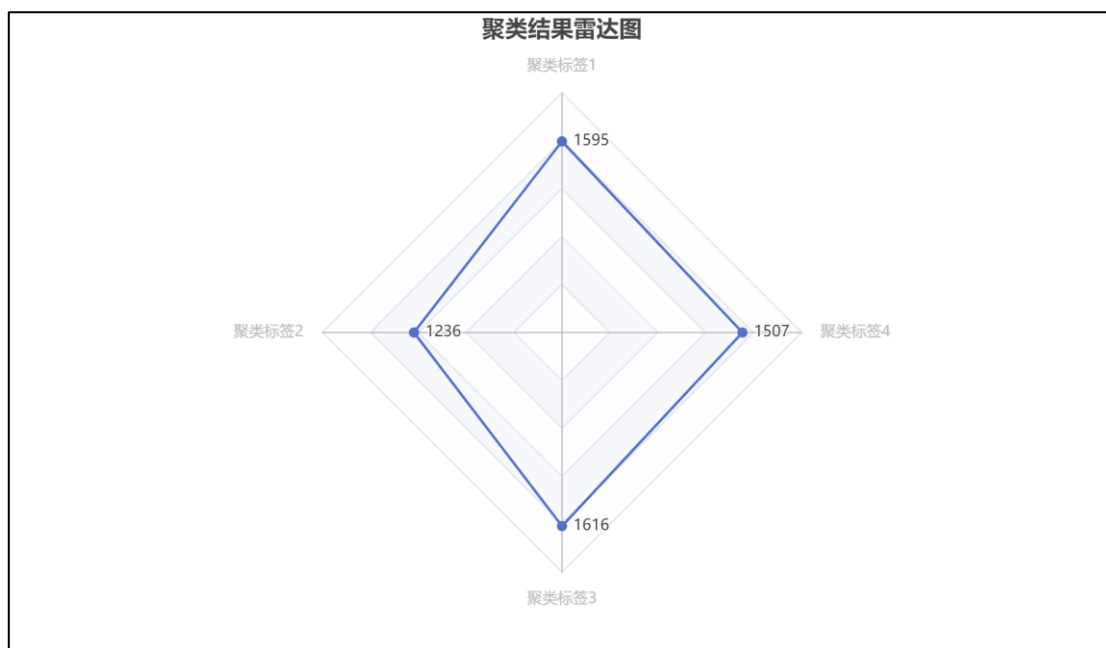


图 2.3.3 雷达图

任务 3 肥料产品的多维度对比分析

3.1 对复混肥料产品登记数量的变化趋势的分析及可视化

分析目标：

从文件“result2_1.xlsx”中提取发证日期中的年份，分析比较复混肥料中各组别不同年份产品登记数量的变化趋势。

3.2 分析广西和湖北两个省份组别的分布差异

分析目标：

从文件“result2_2.xlsx”中提取 2021 年 9 月 30 日仍有效的有机肥料产品，再从有效产品中分别筛选出广西和湖北（根据正式

登记证号区分) 产品登记数量在前 5 的组别, 分析两个省份上述组别的分布差异。

3.3 计算企业之间的杰卡德相似系数矩阵

分析目标:

从附件 3 中提取产品登记数量大于 10 的肥料企业, 给出这些企业所用到的原料集合(发酵菌剂除外)。以各企业用到的原料作为特征, 计算企业之间的杰卡德相似系数矩阵。

任务 4 肥料产品的多维度对比分析

4.1 提取出氮、磷、钾养分和有机质的百分比, 以及肥料含氯的程度

分析目的:

从附件 4 技术指标中提取出氮、磷、钾养分和有机质的百分比, 以及肥料含氯的程度。

4.2 提取各种原料的名称及其百分比

分析目的:

从附件 4 原料与百分比中提取各种原料的名称及其百分比。