

肥料登记数据分析

摘要

我国拥有悠久的农耕文明，广袤的耕地面积以及众多的农业人口，是名副其实的农业大国。肥料作为农作物生长发育的营养元素，是农业生产中的重要生产资料之一，科学合理的使用不同类型的肥料是提高耕地质量的重要措施。现以各肥料销售公司为代表，分析其在农业行政管理部门登记的肥料生产和销售数据，借此分析我国的肥料销售情况及农业发展情况。

目录

1 主要目标

2 任务 1—数据的预处理

2.1 产品通用名称规范化处理

2.1.1 处理思路

2.1.2 处理过程及结果

2.2 总无机养分百分比

2.2.1 处理思路

2.2.2 处理过程及结果

3 任务 2—肥料产品的数据分析

3.1 复混肥料的分组及其特点分析

3.2 有机肥料中无机养分百分比和有机质百分比分组分析

3.3 复混肥料数据的聚类分析

3.3.1 K-means++聚类算法

3.3.2 系统层次聚类法

4、任务三—肥料产品的多维度分析

4.1 复混肥料中各组别登记数量的变化趋势可视化

4.2 广西/湖北两省肥料登记数量分布差异可视化

4.3 不同企业间杰卡德系数计算

5 任务四—算法及处理流程的设计与分析

5.1 技术指标中字符段的提取算法

5.2 原料与百分比中字符段的提取算法

6 结束语

1 主要目标

- 1. 对肥料登记数据作规范化处理
- 2. 以肥料主要养分百分比为基准细分肥料产品
- 3. 实现肥料登记数据多维度对比分析
- 4. 实现非结构化数据的结构化

2 任务 1—数据的预处理

数据挖掘的精度通常受限于获取数据的质量，而我们获取的数据通常是不准确、不完整和不一致的。为了提高数据挖掘的精度，需要对获取的数据进行预处理，数据预处理是指在主要处理之前对数据进行审核、筛选和排序等处理。

2.1 产品通用名称规范化处理

2.2 总无机养分百分比

3 任务 2—肥料产品的数据分析

3.1 复混肥料的分组及其特点分析

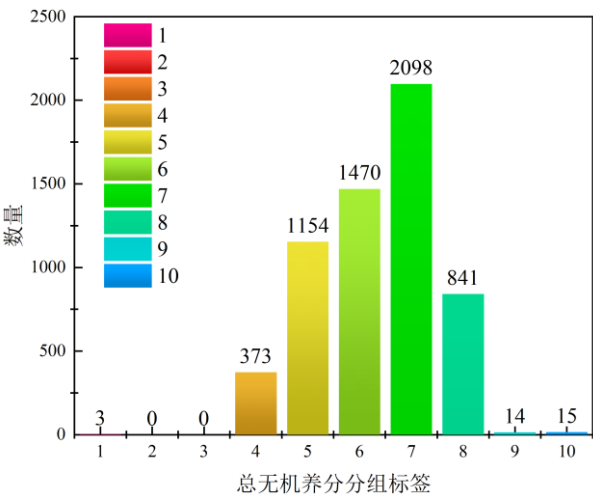


图 5 产品登记数量直方图

排名	一	二	三
分组标签	7	6	5
产品登记数量	2098	1470	1154

3.3 复混肥料数据的聚类分析

4、任务三—肥料产品的多维度分析

4.1 复混肥料中各组别登记数量的变化趋势可视化

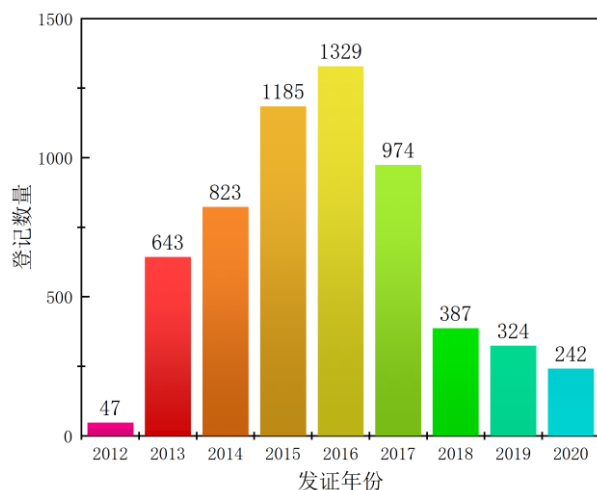


图 13 组别 7 数据展示

4.3 不同企业间杰卡德系数计算

杰卡德相似系数是衡量两个数据几何相似度的一种指标，两个集合交集元素在各自集合中所占比例，其可以由 1 减去杰卡德距离求得。本次杰卡德相似度算法分析，基于 Matlab 平台，运用 PDIST 函数的 JACCARD 类别求得相应数据矩阵的杰卡德距离，进而求得杰卡德相似系数。

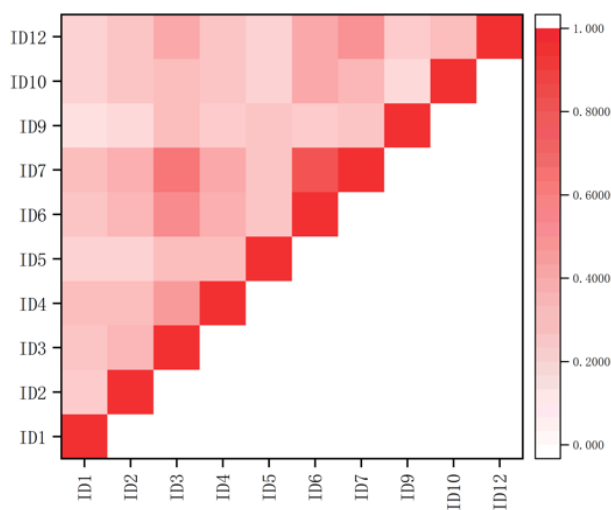


图 16 不同公司产品间的杰卡德相关系数