摘要

肥料是农业生产中一种重要的生产资料，其生产销售，依法在农业行政管理部门进行登记。各省、自治区、直辖市人民政府 ，农业行政主管部门行政区域内的肥料登记数据资源亟待利用。肥料登记数据分析蕴含着前所未有的社会价值和商业价值：为政府决策当智囊，为企业管理做支撑，为学科发展建平台，为管理发展促手段。因此，利用数据分析技术对肥料登记数据的研究具有重大意义。

针对任务一，对数据进行预处理.

1.1产品通用名称存在不规范的情况。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这 4 种类别对其进行规范化处理。

1.2计算各肥料产品的氮、磷、钾养分百分比之和，称为总无机养分百分比。

针对任务二，对肥料产品进行数据分析。

2.1 中筛选出复混肥料的产品，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。根据每个产品所在的分组，为其打上分组标签

2.2 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 (1,1), (1,2),…, (10,10)

针对任务三，对肥料产品进行多维度对比分析

3.1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。在报告中给出处理思路及分析过程，使用合适的图表对结果进行可视化。

3.2 提取 2021 年 9 月 30 日仍有效的有机肥料产品，将完整的结果保存。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

针对任务四，肥料产品的多维度对比分析

4.1设计算法或处理流程，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 设计算法或处理流程，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。

abstract

Fertilizer is an important means of production in agricultural production. Its production and sales shall be registered in the agricultural administrative department according to law. The fertilizer registration data resources in the administrative areas of the people's governments of all provinces, autonomous regions and municipalities directly under the central government and the competent agricultural administrative departments need to be utilized urgently. The analysis of fertilizer registration data contains unprecedented social and commercial value: acting as a think tank for government decision-making, supporting enterprise management, building a platform for Discipline Development and promoting management development. Therefore, using data analysis technology to study fertilizer registration data is of great significance.

For task 1, preprocess the data

1.1 the general name of the product is not standardized. Standardized treatment shall be carried out according to four categories: compound fertilizer (mixed fertilizer is classified into this category), organic-inorganic compound fertilizer, organic fertilizer and bed soil acid conditioner.

1.2 calculate the sum of the percentage of nitrogen, phosphorus and potassium nutrients of each fertilizer product, which is called the percentage of total inorganic nutrients.

For task 2, analyze the data of fertilizer products.

2.1 select the products of compound fertilizer, and divide all compound fertilizers into 10 groups equidistant according to the value of total inorganic nutrient percentage. Label each product according to its grouping

The organic fertilizer products selected in 2.2 are divided into 10 groups equally according to the percentage of total inorganic nutrients and the percentage of organic matter, and each product is labeled with grouping labels (1,1), (1,2),... (10,10)

For task 3, a multi-dimensional comparative analysis of fertilizer products was carried out

3.1 extract the year in the issuing date, analyze and compare the change trend of product registration quantity of each group in compound fertilizer in different years. The processing idea and analysis process are given in the report, and the results are visualized with appropriate charts.

3.2 extract the organic fertilizer products that are still effective on September 30, 2021 and save the complete results. Select the top 5 groups of Guangxi and Hubei (distinguished according to the official registration certificate number) from the effective products, and analyze the distribution differences of the above groups in the two provinces.

For fertilizer enterprises whose registered quantity of extracted products in 3.3 is greater than 10, the set of raw materials used by these enterprises (except fermentation bacteria) shall be given. Taking the raw materials used by each enterprise as the characteristics, calculate the jacquard similarity coefficient matrix between enterprises

For task 4, multi-dimensional comparative analysis of fertilizer products

4.1 design the algorithm or treatment process to extract the percentage of nitrogen, phosphorus, potassium nutrients and organic matter and the chlorine content of fertilizer from the technical indicators in Annex 4. 4.2 design the algorithm or processing flow, and extract the names and percentages of various raw materials from Annex 4 raw materials and percentages.

目录

1、问题分析

1. 分析肥料产品的情况，得出分布特点。
2. 根据肥料产品的分布特点进行多维度对比分析

2、任务一

2.1数据预处理

2.11情况说明

在本次数据分析过程中产品通用名称存在不规范的情况。请按照复混肥料（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这 4 种类别对数据进行规范化处理。具体处理过程将在下文中进行阐述。

2.1数据分析

2.1.1分析情况说明

在本次数据分析过程中，针对已经进行规范化处理之后的数据进行分析。进行以下操作：计算数据中各肥料产品的氮、磷、钾养分百分比之和，称为总无机养分百分比。

3、任务二

3.1.1复混肥料产品的分布特点

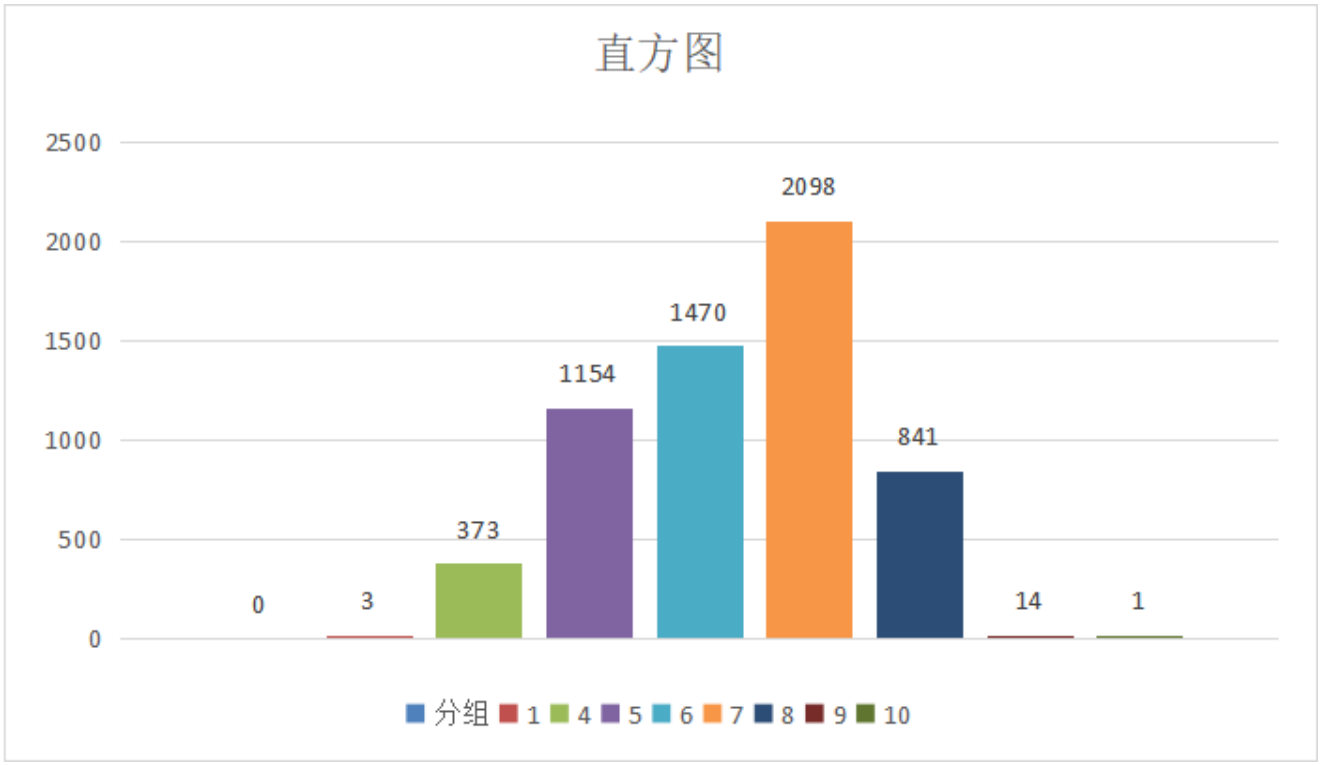
3.1.2复混肥料分组

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 (1,1), (1,2),…, (10,10)

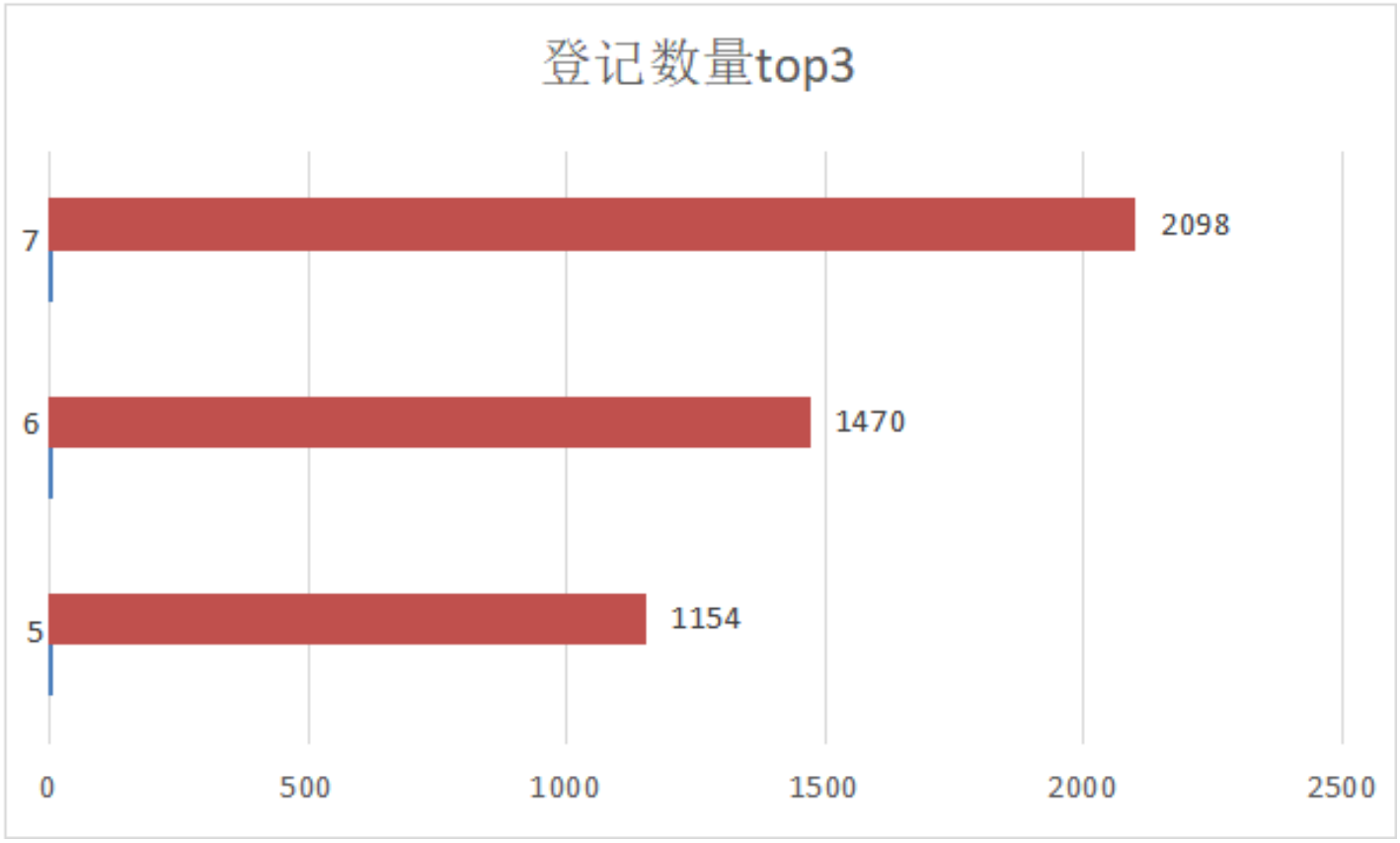
将列名为“总无机养分百分比”中的数值进行排序，再根据排序进行（1-10）组的分组。

区间划分:

```
(-0.072, 7.2]
(21.6, 28.8]
(28.8, 36.0]
(36.0, 43.2]
(43.2, 50.4]
(50.4, 57.6]
(57.6, 64.8]
(64.8, 72.0]
```

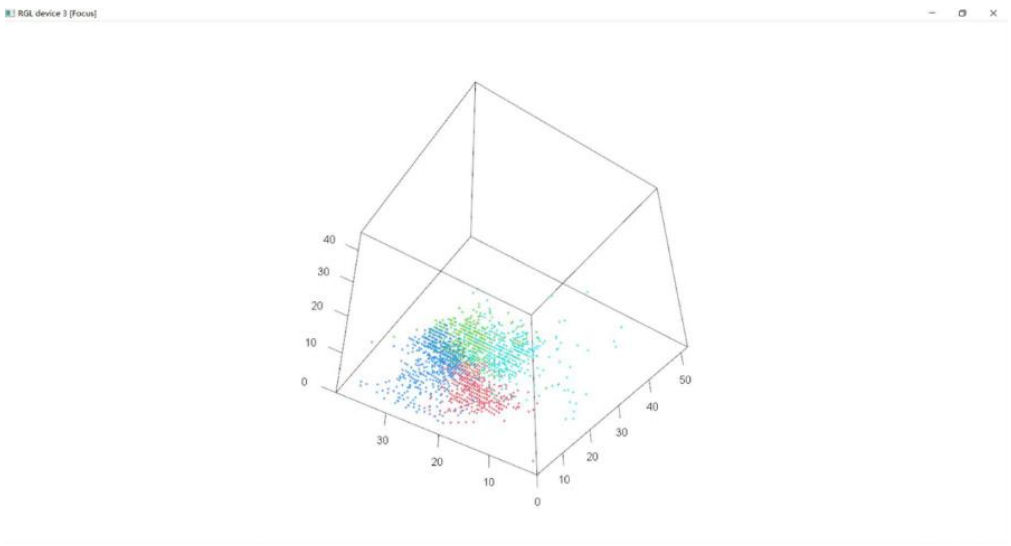


统计登记数量top3

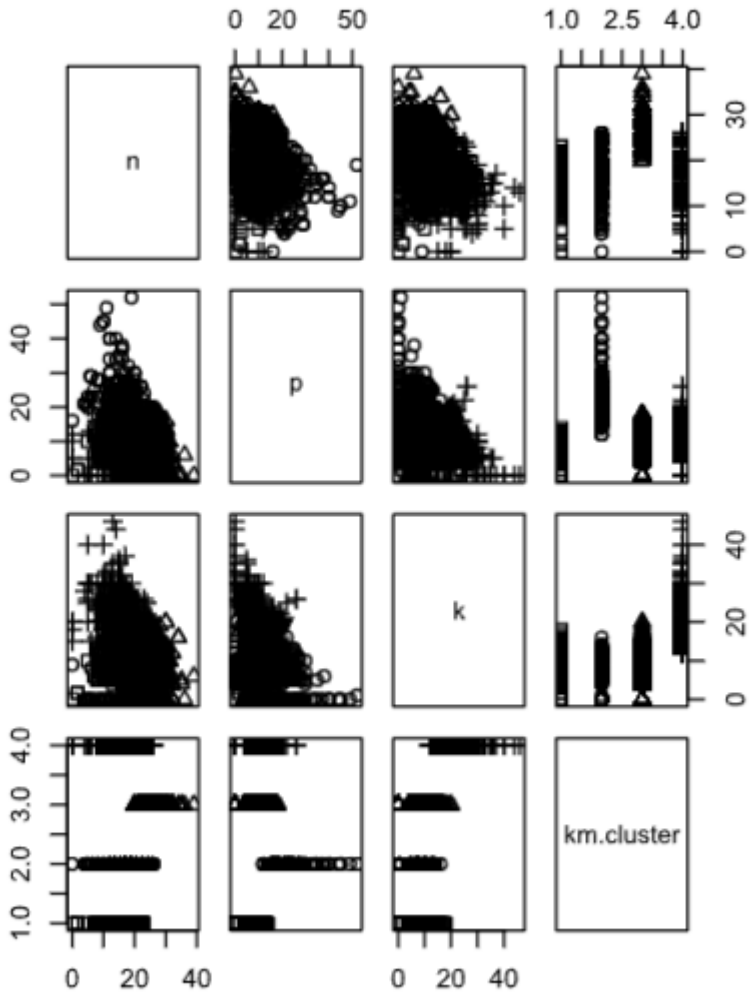


排名	一	二	三
分组标签	7	6	5
产品登记数量	2098	1470	1154

4.1有机肥料产品的分布特点

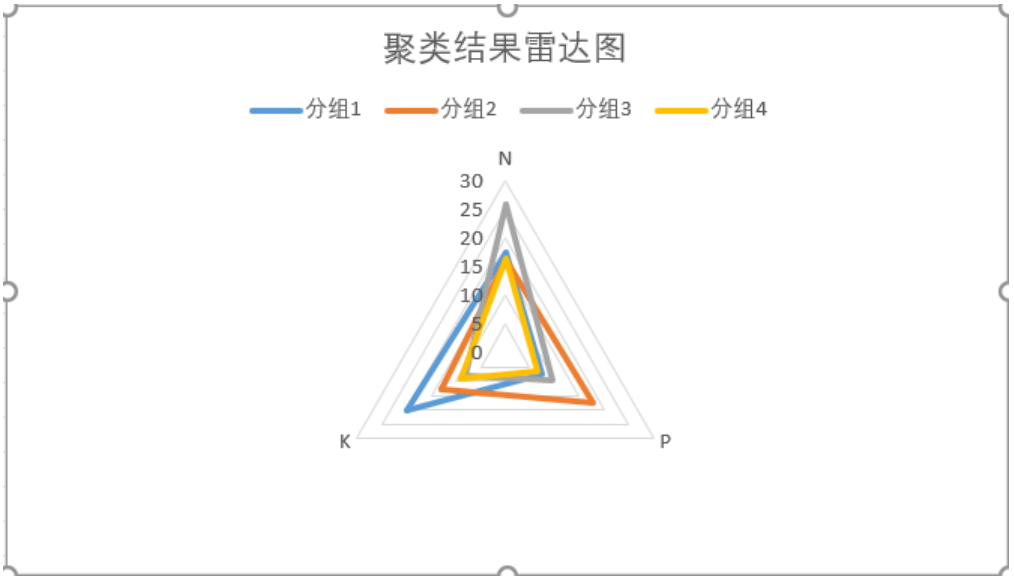


三维散点图



散点图矩阵

雷达图



根据雷达图聚类特征分析：

分组一，钾含量较高，为钾肥组

分组二，磷含量较高，为磷肥组

分组三，氮含量较高，为氮肥组

分组四，钾，磷，氮含量均衡，为均衡组

5.2 提取原料名称百分比

设计算法或处理流程，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。

6、参考文献

[1] 吴达胜,范雪华,姜真杰,曾松伟. 分布式数据挖掘在森林资源信息管理中的应用 [J]. 福建林学院学报. 2004 (04)
[2] 黄兰,孙林山,罗日辉. 数据挖掘及其在图书馆管理中的应用 [J]. 华南热带农业大学学报. 2005 (02)
[3] 张洵. 数据挖掘在农产品安全生产中的应用 [J]. 安徽农业科学. 2007 (34)
[4] 郑斌,李涓子. 农产品市场中数据挖掘的应用 [J]. 安徽农业科学. 2008 (18)
[5] 罗凤娥,彭佳红,胡建强. 数据挖掘在农业领域中的应用研究 [J]. 农业网络信息. 2009 (01)
[6] 梁川,王文生,谢能付. 农业信息资源上数据挖掘的应用 [J]. 中国农学通报. 2009 (11)
[7] 周四方. 数据挖掘在教学中的应用 [J]. 农业网络信息. 2011 (12)
[8] 张超,张娅玲,杨如艳. 数据挖掘在茶叶鉴定中的应用 [J]. 安徽农业科学. 2012 (02)
[9] 许丽莉,贾保先,李寰. Web数据挖掘在农业电子商务中的应用 [J]. 安徽农业科学. 2012 (07)
[10] 梁第,曲延庆,张铭丽. 数据挖掘理论算法综述 [J]. 山东省农业管理干

部学院学报. 2012 (05)