

项目结题验收单

专家验收表（主持人所在单位组织 3-5 名专家对项目进行验收、自评。）

项目名称	农林院校图书馆对科技小院数据资源服务的保障路径研究——以中国农业大学为例			
主持人	王珂	职务/职称	助理馆员	
所在单位	（加盖单位公章）中国农业大学图书馆			
专 家 意 见	<div style="text-align: center;">  </div> <p>本研究聚焦农林院校图书馆对科技小院的数据资源服务保障问题，以中国农业大学科技小院的工作日志为基础，通过文本挖掘、TF-IDF 算法及 HDBSCAN 密度聚类等方法，分析了科技小院的信息需求特征，构建了“数据清洗-主题聚类-需求识别”分析框架，并提出了五维协同保障路径及优化方向。</p> <p>研究以科技小院实践日志为数据载体，系统梳理了农业生态安全、生产技术应用等核心需求的分布特征，并针对需求层级提出了图书馆数据资源服务的具体保障措施。研究过程规范，分析结果对优化图书馆资源配置、提升服务精准性有一定参考作用。建议后续可进一步扩展样本覆盖范围，纳入多区域、多类型机构的科技小院日志，补充实验报告、采访记录等非结构化资料，更全面反映科研生活实际需求；优化主题聚类方法，提升语义解析的准确性。同意结题。</p> <p style="text-align: right;">（如需要可增加页数）</p>			
专家签字	徐伟	于莹	周丽英	
职务/职称	副研究员	馆员	副研究馆员	



项目编号:

注:项目编号请查看立项
通知,也可缺省

CALIS 全国农学文献信息中心研究项目 结题报告

项目名称: 农林院校图书馆对科技小院数据资源服务的保障路径研究——以中国农业大学为例

项目关键词: 科技小院; 数据资源; 图书馆; 乡村振兴

项目单位(盖章): 中国农业大学图书馆

通信地址:(详细地址含邮编) 北京市海淀区清华东路 17 号中国农业大学东校区 100083

项目主持人: 王珂

联系电话: 010-62738640

电子邮件: wangke@cau.edu.cn

提交日期: 2025 年 5 月 30 日

农林院校图书馆对科技小院数据资源服务的保障路径研究

——以中国农业大学为例

目录

1. 研究背景和意义	2
1.1 研究背景	2
1.2 研究意义	4
2. 研究方法	4
2.1 文献研究法	5
2.2 文本挖掘法	5
2.3 聚类分析法	5
3. 研究内容	6
3.1 数据来源	6
3.2 数据处理	6
3.3 文本特征分析	7
3.4 主题聚类分析	10
3.5 保障路径分析	18
4. 结论与展望	26
4.1 研究结论	26
4.2 研究展望	26
5. 参考文献	27
6. 附录：停用词表	28
7. 研究成果	28

农林院校图书馆对科技小院数据资源服务的保障路径研究

——以中国农业大学为例

王珂 贺玢 杨学红 雒庆生
(中国农业大学图书馆 100083)

摘要：2025年“推广科技小院模式”再次被写入中央一号文件。科技小院作为农业科技创新与推广的重要载体，在科研实践、技术示范和社会服务以及乡村振兴中起着重要作用，其对信息资源有着迫切的需求。然而，当前高校图书馆对科技小院的信息服务缺乏系统性与针对性。为精准把握师生的资源服务需求，本研究以中国农业大学28个科技小院的1906篇工作日志为研究对象，采用文本挖掘TF-IDF算法与HDBSCAN密度聚类分析法深入挖掘工作日志并创新性地构建了“数据清洗-主题聚类-需求识别”分析框架，揭示科技小院五维需求主题的层级分布特征，其中农业生态安全与科研体系、生产技术应用、经济全球化协同领域信息需求较为突出，而政策制度与智能系统信息需求相对较弱。基于此，提出对应农业经济支撑、技术社会化应用支持、政策民生配套、生态科研协同及智能知识工程赋能服务的五维需求的协同保障路径，并结合“政产学研用”融通模式，从需求分层响应、文化科技融合、制度长效保障及场域价值提升四个方面提出优化方向，提升农林高校图书馆对科技小院的服务能力和水平，助力科技小院长足发展，从而为实现乡村全面振兴贡献力量。

关键词：科技小院；数据资源；图书馆；乡村振兴

1. 研究背景和意义

1.1 研究背景

1.1.1 科技小院的信息服务需求

(1) 科技小院简介

科技小院（Science and Technology Backyard）最早于2009年由中国农业大学资源与环境学院主办成立，位于河北曲周县白寨乡，以实现作物高产和资源高效（简称“双高”）为目标，以零距离、零门槛、零时差和零费用服务为手段，以研究生与科技人员长驻为特色的集农业科技创新、示范推广和人才培养于一体的科技服务平台^[1]。截至2024年5月17日，全国已建立科技小院1800余个，覆盖31个省、自治区、直辖市，与全世界分享中国的乡村振兴经验。其中中国农业大学牵头的科技小院官方数量为241个^[2]。2009至2024年的15年来，科技小院累计服务近600个贫困村、4.8万个贫困户、17.7万贫困人口，2023年荣获第三届全国创新争先奖，2024—2025年“推广科技小院模式”两次被写入中央一号文件。科技小院还走出了国门，推广到老挝和非洲8国，被联合国粮农组织作为重点案例向全球推广应用^[3]。

(2) 科技小院作用

创新人才培养模式：科技小院是中国农业大学探索研究培养模式改革的重要成果，打破了传统农业人才培养的局限，将理论学习与实践应用紧密结合，使研究生在解决实际问题中提升专业能力和综合素质^[4]。在科技小院模式下，每位研究生每年至少驻扎 120 天，撰写工作日志不少于 100 篇。这种长期驻扎和深入实践，使研究生能够真正了解农民需求，解决农业生产中的实际问题。

科研与社会服务相结合：科技小院将科学研究与社会服务融为一体，形成了“顶天立地”的科研模式。一方面，研究生在农业生产一线开展科学研究，解决实际问题；另一方面，他们将科研成果转化为农民可用的技术，直接服务农业生产^[5]。这种模式使中国农业大学的科研工作更加贴近农业生产实际，提高了科研成果的应用价值和社会效益。同时，通过科技小院，学校将先进的农业技术和理念带到农村，促进了农业现代化和乡村振兴。

服务乡村振兴战略：科技小院是中国农业大学服务乡村振兴战略的重要抓手。通过科技小院，学校将人才、技术、信息等资源带到农村，帮助农民解决生产中的实际问题，提高农业生产效率和农产品质量。2024 年中央一号文件明确提出支持科技小院扎根农村助农惠农，这充分体现了国家对科技小院模式的认可和支持^[6]。科技小院已成为中国农业大学服务乡村振兴战略的重要平台和品牌^[7]。

(3) 信息服务需求

尽管科技小院在乡村振兴中展现出显著效能，其规模化扩张与全球化布局却逐渐暴露出支撑体系的深层次矛盾。当前，全国逾 1800 个科技小院每年产生大量的田间实验数据，但由于缺乏统一的数据整合平台，许多数据仍分散存储于各地方站点，形成“数据碎片化”现象。这不仅导致大量科研成果难以跨区域复用，更使得技术协作周期延长。

同时，人才培养的结构性矛盾日益凸显，部分驻院研究生对智能农业装备操作、农业大数据分析等技能掌握不足，实践中过度依赖传统农技经验，难以满足数字乡村建设对“技术+管理+数据”复合型人才的需求。此外，数据资源的存储、分析及决策支持能力薄弱，使得部分小院在应对区域性病虫害预警、气候适应性种植等复杂问题时捉襟见肘。这些短板表明，亟需通过资源整合机制创新与人才培养模式重构，为科技小院注入持续发展动能。

1.1.2 高校图书馆面向科技小院的信息服 务挑战

(1) 高校图书馆对科技小院服务

高校图书馆是科技小院研究生获取知识资源的重要渠道。图书馆提供的信息服务对于支持科技小院研究生的学习和研究具有重要意义。例如，中国农业大学图书馆为全校科技小院的同学们提供图书资料、电子资源获取等情报信息服务，并面向同学们征集在这些方面的意见，以改进服务。

(2) 高校图书馆信息服务的挑战

目前，中国农业大学图书馆对科技小院的服务支撑已从基础保障向精准赋能演进。以中

国农业大学图书馆 2023 年成立的“科技小院服务站”为例，其深入了解科技小院的资源需要，主动提供各项服务，找到高校图书馆为科技小院提供高质量资源之路，优化资源配置。虽然做了以上尝试与改进，但图书馆对科技小院的信息服务仍存在诸多挑战：

数字鸿沟：有些科技小院所在农村地区缺乏可靠的互联网连接，阻碍科技小院学生访问数字资源。虽然图书馆提供远程访问，但某些资源由于许可限制等原因在某些地区无法使用。

个性化支持：一方面，科技小院的学生存在不完全了解图书馆提供的所有资源和服务；另一方面，图书馆工作人员难以向位于不同地点的科技小院学生提供面对面的个性化支持，特别是使用图书馆的资源服务的相关培训。

资源不足：图书馆的资源不可能涵盖科技小院学生在农村地区特定农业问题。向各个科技小院学生配送实体资源（如书籍）可能会延迟，影响研究学习。

反馈机制：图书馆缺乏有效的机制来收集和实施科技小院学生对资源与服务的反馈，导致资源服务与学生需求不匹配。

因此，本文聚焦科技小院学生们的工作日志，对文本内容进行深入分析与挖掘，精准把握师生的资源需求，构建针对科技小院师生的数据资源服务保障路径模型，提供系统性、有针对性地资源保障与科研支撑。

1.2 研究意义

研究意义可分为理论、实践、政策响应三个层面。

1.2.1 理论层面

通过解析科技小院工作日志，揭示科技小院一线科研与生产、生活场景下的资源需求及其类别。在此基础上，提出了科技小院数据资源服务保障路径并进行优化，突破传统图书馆资源服务路径依赖，为涉农院校知识服务模式创新提供理论依据。

1.2.2 实践层面

提升图书馆资源服务精准性与可持续性。具体体现在提出的保障路径可以优化资源配置效率，建立对应政府、科研团队与农民各主体服务的分层响应机制；构建农业知识体系，缩短技术转化链条、缩短实验室成果到田间应用的周期，且增强服务可复制性；制定规范制度保障科技小院模式长效发展等方面。

1.2.3 政策响应层面

探索高校服务乡村振兴的实施路径。在 2024、2025 年中央一号文件“强化农业农村科技服务”的政策导向下，研究通过实证检验“高校图书馆—科技小院”协同服务模式，为落实“推广科技小院模式”提供可操作的资源支撑方案，助力实现“将论文成果写在乡村大地”的实践要求。

2. 研究方法

为了深入探讨农林院校图书馆对“科技小院”数据资源服务保障路径，研究采用文献研究法、文本挖掘法、聚类分析法三种方法进行研究。并采用图 1 所示技术路线图实施。

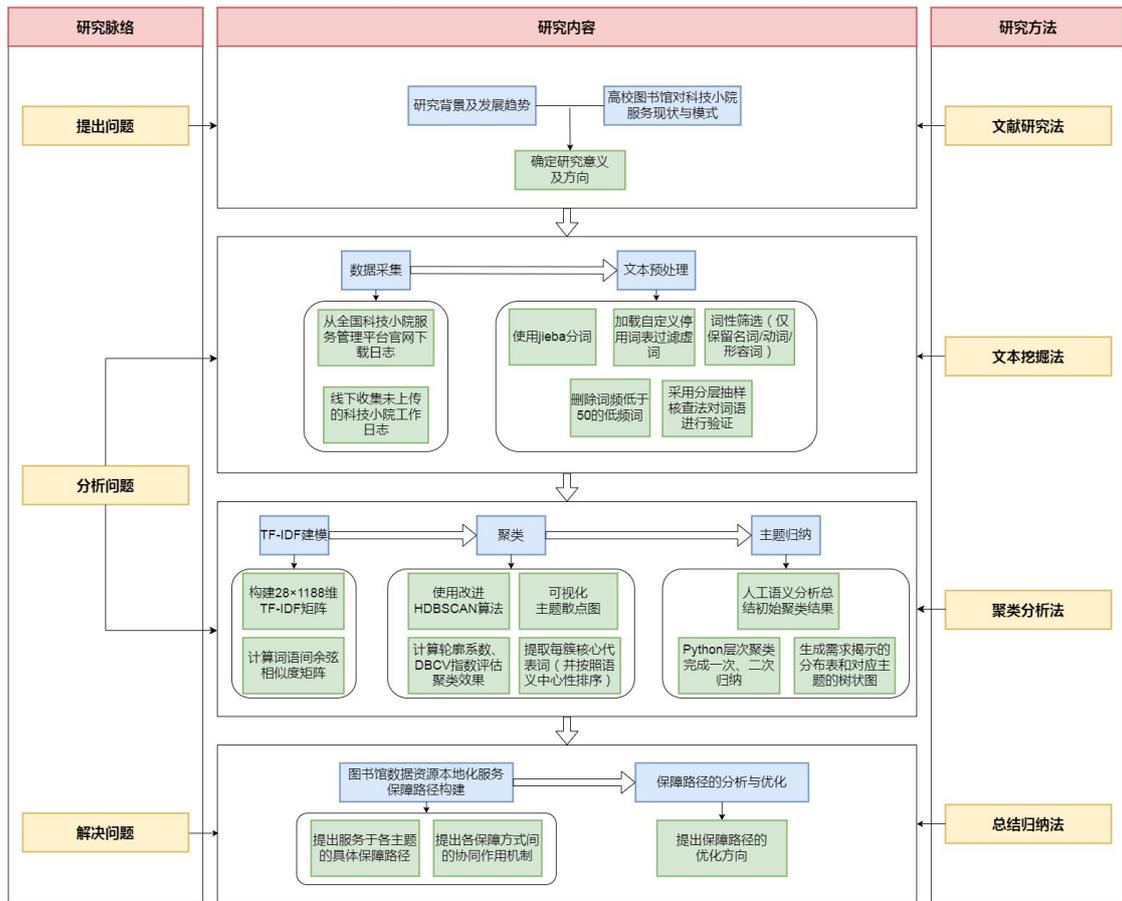


图 1 技术路线图

2.1 文献研究法

文献研究法是一种通过收集、分析、综合和评价已有文献资料来深入研究某个问题或主题的方法。该方法包括确定研究主题或问题、收集文献资料、筛选文献、阅读和分析文献、综合和评价文献，以及撰写文献综述等步骤^[8]。通过相关文献调研深入了解科技小院的产生背景与发展趋势，高校图书馆对科技小院的信息服务现状与模式，明晰研究背景与意义，最终确定研究目标与研究方向。

2.2 文本挖掘法

文本挖掘是从文本数据中发现隐藏的、有价值的信息和知识的过程。它结合了自然语言处理（NLP）、机器学习、数据挖掘、统计学和信息检索等多学科领域的技术，通过对文本内容的分析和挖掘，帮助用户更好地理解 and 利用文本数据^[9]。本研究采用 Jieba 分词技术对科技小院日志进行语义解构与需求识别，再通过实证分析，揭示农林院校图书馆资源服务的优化方向。

2.3 聚类分析法

聚类分析是一种将数据集中的对象划分为若干个子集（或称为簇）的分析方法，使得同一簇内的对象之间相似度较高，而不同簇之间的对象相似度较低。聚类分析是一种无监督学

习方法，即在聚类过程中不需要预先定义类别的标签，而是根据数据本身的特征和相似性自动发现数据的内在结构^[10]。本文采用 TF-IDF 算法将文本转换为数值型向量从而完成文本数据分类。并结合 UMAP 和 HDBSCAN 算法实现更高效和更准确的数据聚类分析。

3. 研究内容

3.1 数据来源

采用中国农业大学科技小院的电子版工作日志作为研究数据。工作日志的来源为“全国科技小院服务管理平台”下载及线下联系科技小院相关老师获得。

截至 2024 年 5 月 17 日最终搜集到电子版工作日志共 1906 篇，涉及科技小院数量为 28 个，具体小院名称及日志数量如图 2 所示。

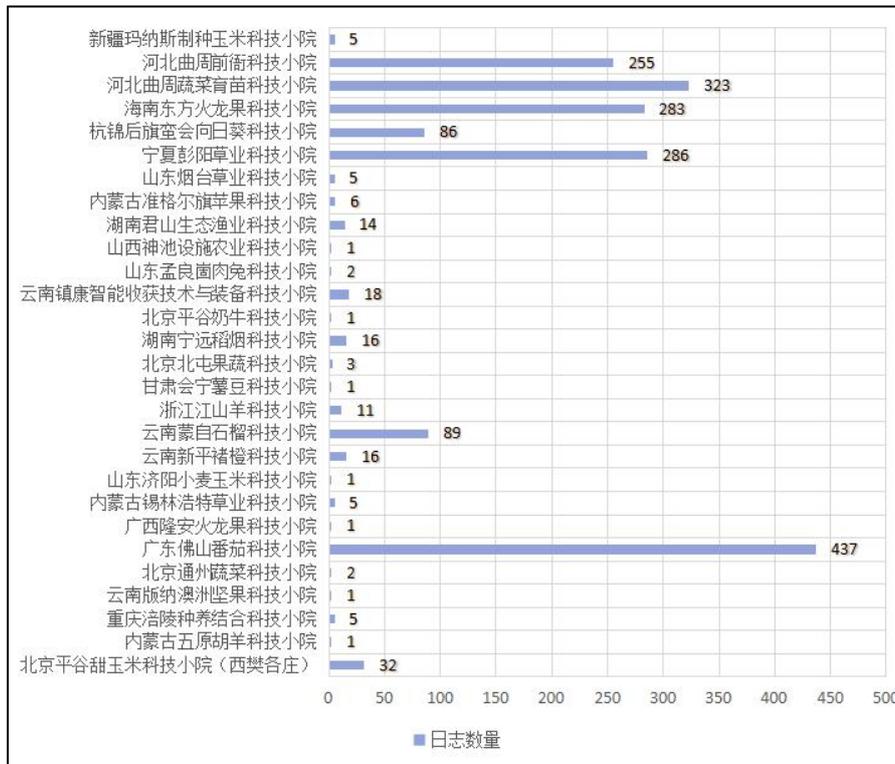


图 2 科技小院电子版工作日志数量

3.2 数据处理

首先，对日志文本数据清洗。包括去除所有的标点符号以及多余的空格，处理后得到大约 205 万字的纯文本数据。再利用 Python 和 jieba 对文本数据进行初步切词，并基于结果编制一份停用词列表（见附录）。停用词定义为经常出现在文本中但对信息检索没有帮助的应该消除的词语^[11]，如“的、是、太”等等^[12]。同时，在文本挖掘中，停用词的判断更侧重于其是否能够表示文本特征^[13]。

第二步，对日志文本词性标注（Part-Of-Speech tagging，简称 POS）。选择性地仅保留名词、动词和形容词这三种词性的词，切词数量为 32898 个。词性作为词语基本的语法属性，是词语和语句的关键性特征^[14]，名词通常代表实体和概念，动词描述了行为和过程，而形容词则用于修饰名词，增强语言的表现力。因此，通过聚焦于这些关键词可以更有效地分析

日志内容，揭示科技小院中相关活动的特征和趋势。

第三步，对日志文本词频统计（见图3）。可知词语出现次数大于等于10次的词汇有6275个，词语出现次数大于等于50次的词汇有1950个，占总词汇数量32898的5.9%。根据自定义阈值确定法^[15]，初步确定研究的高频词阈值为50次，即仅选取词语出现次数大于或等于50次的词汇作为分析对象。

第四步，为系统评估特征词筛选结果的代表性，本部分采用分层抽样核查法实施语义效度检验。从高频词库（ $n \geq 50$ ）中抽取30个特征词进行人工语义验证，覆盖名词、动词、形容词三类词性。核查发现，绝大多数词汇直接关联农业生产技术（如“施肥”“病虫害”）或资源需求场景（如“文献”“检索”），个别词汇（如“处理”“事情”）需结合上下文语境进一步研判其语义贡献度。该过程证实特征词筛选大部分能有效捕捉科技小院科研人员需求的核心语义。



图3 词频数量统计

为了进一步确定文本中意义不明的语义，如“进行”，无法判断具体“进行”的内容，再对词汇进行一次筛选，筛选后的词汇数量为1188个（见表1）。由此可知，出现次数越多的词汇数量越少，频次在50-100的词汇最多，为41.7%，频次在1000以上的占3.5%。

表1 词语出现次数数量情况

	$n \geq 1000$	$500 \leq n < 1000$	$200 \leq n < 500$	$100 \leq n < 200$	频次 ≥ 50 词汇个数
词汇个数	41	94	243	315	1188
占大于50次的总数比	3.5%	7.9%	20.5%	26.5%	100.0%

根据 Adelman 等人提出的“可能需要原则”^[16]，某一词汇的出现次数越多，往往意味着它在多种语境下都可能被需要，从而反映了其背后所反映需求的重要性。但词语的出现频次并不能完全体现其重要性，故进一步使用 TF-IDF 算法对词语与文本进行实证分析量化词语权重特征。

3.3 文本特征分析

3.3.1 TF-IDF 矩阵构建

选取自然语言处理（natural language processing, NLP）领域的 TF-IDF 算法（Term Frequency-Inverse Document Frequency），TF-IDF 算法是由 Salton 等^[17]提出的一种应用于挖掘数据和检索信息的加权技术，是目前使用频率最高的特征权重算法，可以量化词语在语料库中的权重特征^[18]。基于词频统计的 TF-IDF 算法是一种无监督抽取算法，无需昂贵的人工

标注，更适合大规模文本数据的快速处理和分析^[19]。

TF-IDF 算法在本研究中用于衡量词语在文档集中的重要性，具体为 1188 个词语及它们在 28 篇文档中的情况。

构建 TF-IDF 矩阵前，需要分别计算每个词语的 TF（Term Frequency，词频）值和 IDF（Inverse Document Frequency，逆文档频率）值。TF（词频）是衡量一个词语在文档中出现的频率。它通常被计算为词语在文档中出现的次数除以文档的总词数，计算公式如下：

$$tf_i = \frac{n_{ij}}{\sum_k n_{kj}} \quad (1)$$

其中：式（1）表示在第 j 个文档中计算词的词频 tf_i ； n_{ij} 表示第 j 个文档中第 i 个词出现的次数； n_{kj} 表示第 j 个文档中第 k 个词语出现的次数， $\sum_k n_{kj}$ 表示第 j 个文档中所有词出现次数的总和。图 4 是词语在文档中 TF 值的统计情况表（部分）。

Words	F_rizhi00	TF_rizhi002	TF_rizhi003	TF_rizhi004
工作	0.11031895	0.02259887	0	0.026217228
小院	0.14258912	0.02259887	0	0.05093633
农业	0.09193246	0	0	0.013483146
生长	0.01575985	0.002824859	0	0.006741573
种植	0.07166979	0.002824859	0	0.019475655
甜叶菊	0	0	0	0
试验	0.07917448	0	0.013513514	0.002247191
土壤	0.04727955	0.002824859	0	0.005243446
问题	0.05103189	0.002824859	0	0.019475655
处理	0.01313321	0	0	0.007490637
老师	0.01688555	0.002824859	0	0.009737828
发展	0.04990619	0	0	0.008988764
时间	0.01313321	0.016949153	0	0.008988764
科技	0.08142589	0.008474576	0	0.01423221
影响	0.03639775	0.005649718	0.054054054	0.002247191
准备	0.0097561	0.005649718	0	0.012734082
技术	0.10131332	0	0	0.001498127
实验	0.0108818	0.008474576	0	0.011235955
生产	0.04915572	0	0	0.011985019
学习	0.02251407	0	0	0.019475655

图 4 词语在文档中 TF 值的统计情况表（部分）

IDF（逆文档频率）用于衡量一个词语在语料库中的普遍程度。它通常被计算为语料库中文档总数的对数除以包含该词语的文档数。IDF 值越高，说明该词语在语料库中的区分度越高，计算公式如下：

$$idf_i = \lg \frac{|D|}{|\{j:t_i \in d\}|} \quad (2)$$

式（2）用以计算逆文档频率 idf_i 。其中： D 是所有文档的数量总和（本研究为 28，代表 28 个小院的日志）； $|\{j:t_i \in d\}|$ 是整个文档集 d 中包含文档中词语 t_i 的文档数量； $\{j:t_i \in d\}$ 是包含词 i 的文档数量。图 5 是各词语 IDF 值的统计情况（部分）。

Words	IDF
工作	1.1892
小院	1.4229
农业	1.3228
生长	1.5341
种植	1.4769
甜叶菊	3.2687
试验	1.5947
土壤	1.3716
问题	1.3228
处理	1.5947
老师	1.4769
发展	1.2763

图 5 各词语的 IDF 值（部分）

TF-IDF 矩阵是将 TF 和 IDF 相结合，得到每个词语在文档中的 TF-IDF 权重。这个权重既考虑了词语在文档中的重要性(通过 TF)，也考虑了词语在语料库中的区分度(通过 IDF)，具体计算公式如下：

$$TF - IDF_i = tf_i \times idf_i \quad (3)$$

即将 TF 与 IDF 值求积。图 6 是各词语在各文档中 TF-IDF 值的统计情况表（部分）。

Words	TF-IDF_rizhi001	TF-IDF_rizhi002	TF-IDF_rizhi003	TF-IDF_rizhi004
工作	0.131195928	0.026875525	0	0.031178629
小院	0.202883904	0.032154957	0	0.072475106
农业	0.121605809	0	0	0.017835147
生长	0.02417691	0.004333566	0	0.010342129
种植	0.105850843	0.004172102	0	0.028764064
甜叶菊	0	0	0	0
试验	0.126260112	0	0.021550096	0.003583611
土壤	0.064846907	0.003874473	0	0.007191719
问题	0.067503633	0.003736648	0	0.025761879
处理	0.020943621	0	0	0.011945372
老师	0.02493868	0.004172102	0	0.014382032
发展	0.063692945	0	0	0.01147194
时间	0.019396751	0.025032611	0	0.013275722
科技	0.103920069	0.010815707	0	0.018163906
影响	0.044834805	0.006959331	0.066583871	0.002768094
准备	0.014966658	0.008667133	0	0.019535133
技术	0.138957659	0	0	0.002054777
实验	0.016693581	0.013000699	0	0.017236882
生产	0.069941556	0	0	0.017052966
学习	0.032034301	0	0	0.02771107

图 6 词语在文档中 TF-IDF 值的统计情况表（部分）

TF-IDF 矩阵这个矩阵的每一行代表一个文档，每一列代表一个词语，矩阵中的值是该词语在对应文档中的 TF-IDF 权重。

鉴于 TF-IDF 模型依赖 IDF 对局部高频词的降权效果，其数学特性需进行验证，确保 TF-IDF 算法对局部高频词语义偏差的有效抑制，尤其针对仅集中在少数文档的高频词。本研究进一步实施跨文档分布均衡性检验，以验证特征词的全局代表性。通过观察词语在 28 个小院文档中的出现频次分布，发现高频词普遍呈现跨小院覆盖特性，即同一词汇在多数小院日志中均有稳定出现，而仅在小部分文档中异常高频的词汇已被 TF-IDF 算法过滤。例如，“种植”等高频词在 20 个以上小院中重复出现，体现农林技术服务的共性需求。以上跨文档分布均衡性检验，以及文本挖掘中对 $N \geq 50$ 次的词语进行人工语义验证的结果，均说明当前文本处理方式能够平衡词汇的局部重要性（词频）与全局区分度（逆文档频率），从而规避“局部高频词伪代表性”问题。

基于此通过 Python 对小院日志集使用 TF-IDF 算法进行处理。接下来再计算余弦相似度矩阵判断词语间的相关程度，再使用 HDBSCAN 密度聚类方式完成词语聚类，并获取聚类后各类词语组中的关键代表词。

3.3.2 余弦相似度矩阵

余弦相似度是一种广泛应用于自然语言处理和文本挖掘领域的相似度度量方法，它通过计算两个向量之间夹角的余弦值，消除文本长度干扰，评估词语语义的相似程度^[20,21]。余弦值越接近 1，表示两个向量的方向越一致，即词语之间的语义相似度越高；反之，余弦值越接近-1，则表示词语之间的语义差异越大^[22]。根据余弦相似度的值，可以对文档进行聚类、分类或排序。

具体计算步骤参考 Ped regosa 等（2020）^[23]提出的数值优化方案，使用 Python 中 scikit-learn 库的 cosine_similarity 函数进行批量计算，该实现方式通过矩阵运算优化提升了计算效率。

$$\cos(d_i, d_j) = \frac{(d_i \cdot d_j)}{\|d_i\| \times \|d_j\|} \quad (4)$$

公式(4)为余弦相似度计算公式,其中 d_i 和 d_j 分别表示两个词语向量。生成的余弦相似度矩阵(部分)如图7所示:

	A	B	C	D	E	F	G	H	I	J	K	L
1		工作	小院	农业	生长	种植	甜叶菊	试验	土壤	问题	处理	老师
2	工作	1	0.565193	0.691943	0.436503	0.627737	0.235369	0.752204	0.554155	0.808635	0.464493	0.358598
3	小院	0.565193	1	0.659664	0.197676	0.630996	0.292955	0.465615	0.567033	0.494336	0.400525	0.632388
4	农业	0.691943	0.659664	1	0.491889	0.843573	0.428849	0.518583	0.615127	0.640603	0.478708	0.429974
5	生长	0.436503	0.197676	0.491889	1	0.733818	0.149799	0.244683	0.698657	0.342504	0.63245	0.277466
6	种植	0.627737	0.630996	0.843573	0.733818	1	0.417694	0.481619	0.745336	0.562705	0.595747	0.465526
7	甜叶菊	0.235369	0.292955	0.428849	0.149799	0.417694	1	0.112294	0.213255	0.228226	0.294094	0.367274
8	试验	0.752204	0.465615	0.518583	0.244683	0.481619	0.112294	1	0.582469	0.644822	0.552613	0.465709
9	土壤	0.554155	0.567033	0.615127	0.698657	0.745336	0.213255	0.582469	1	0.502862	0.731166	0.698434
10	问题	0.808635	0.494336	0.640603	0.342504	0.562705	0.228226	0.644822	0.502862	1	0.580266	0.380764
11	处理	0.464493	0.400525	0.478708	0.63245	0.595747	0.294094	0.552613	0.731166	0.580266	1	0.609687
12	老师	0.358598	0.632388	0.429974	0.277466	0.465526	0.367274	0.465709	0.698434	0.380764	0.609687	1

图7 余弦相似度矩阵(部分)

为了验证以上各矩阵中数值的准确性,采取抽样检验的方法,随机挑选三个词语,并针对这些词语分别计算了它们的TF值、IDF值以及TF-IDF值进行校验。此外还随机选定了—个词语,计算了它与前述三个词语之间的余弦相似度。经过—系列的检验,所得结果均准确无误。

3.4 主题聚类分析

在计算出词语间的余弦相似度矩阵之后,再进行词语聚类,并从各个聚类中提炼出具有代表性的关键词。词语聚类是一种无监督的学习方法,它的基本原则是使类内的相似度尽可能大,而类间的相似度尽可能小^[24]。

研究采用改进的HDBSCAN算法体系。该算法广泛应用于数据挖掘、图像分割、文本聚类等领域,尤其适合探索性数据分析和异常检测,有以下3个优点:

- (1) 层次性(Hierarchical):通过构建密度层次结构,自动合并或拆分不同密度的聚类,无需预先指定聚类数量。
- (2) 基于密度(Density-Based):通过分析数据点的密度分布来识别聚类,能有效处理不规则形状的簇和噪声数据。
- (3) 支持噪声处理(With Noise):明确区分噪声点(离群点)与核心聚类区域,增强结果的鲁棒性。

在使用HDBSCAN算法的基础上结合相似度矩阵增强、UMAP降维和分层聚类策略,构建包含数据预处理、特征优化、层次聚类和结果解释的四阶段处理流程^[25]。通过动态参数调整和二次聚类机制,解决传统密度聚类在语义分析中的噪声敏感性和簇形态适应性问题。

3.4.1 相似度矩阵增强

在输入词语的余弦相似度矩阵后,执行矩阵优化处理。首先对矩阵完成对称化修正消除计算偏差,再进行高斯增强两步以增强词语间的局部相似性,公式为:

$$S'_{i,j} = \exp\left(-\frac{(1-S_{i,j})^2}{2\alpha^2}\right) \quad (5)$$

其中调节因子 $\alpha = 0.3$,可以在强化近邻关系的同时保持全局结构。

3.4.2 降维空间映射

采用 UMAP 算法将高维语义空间投影至二维可视化空间，参数配置如下：邻域规模 $n_neighbors=20$ ，可平衡局部/全局结构保留；最小间距 $min_dist=0.1$ ，避免簇间过度重叠；距离度量取 cosine 相似度，契合语义特征空间特性。

3.4.3 层次化密度聚类（HDBSCAN）

在主聚类阶段构建最小生成树识别核心语义簇，控制参数确定簇合并阈值；当主簇规模超过 50 个元素时，启动二次聚类。参数可动态调整实现细粒度语义划分。图 8 展示部分最终分组情况。

词语	主聚类ID	最终分组
工作	7	-1
小院	7	27
农业	6	6
生长	7	8
种植	7	20
甜叶菊	1	1
试验	7	-1
土壤	7	24
问题	7	-1
处理	7	23
老师	7	18
发展	6	6
时间	7	-1
科技	7	-1
影响	7	24
准备	7	24
技术	7	12

图 8 分组情况（部分）

3.4.4 聚类结果及评估

采用轮廓系数（Silhouette Coefficient）^[26]和 DBCV 指数（Density-Based Cluster Validity，专门评估密度聚类质量）^[27]定量评估聚类效果，同时将聚类结果生成散点图观察聚类效果。轮廓系数是一种常用的评估聚类分析效果的指标，可用于大多数聚类方法，其取值范围为[-1, 1]，轮廓系数接近 1 时表明聚类效果非常好，意味着大多数对象与其自身的簇内其他成员非常相似，轮廓系数接近 -1，则表明聚类效果较差^[28]。因此，为了获得良好的聚类结果，应追求较高的正向轮廓系数值。公式计算如下：

$$Silhouette(i) = \frac{b(i)-a(i)}{\max[a(i),b(i)]} \quad (6)$$

其中 $a(i)$ 为簇内相似度， $b(i)$ 为最近邻簇相似度。通过过滤噪声点（label=-1）计算有效聚类的整体轮廓系数。见图 14，本研究计算轮廓系数为 0.529，表示聚类效果较合理。

DBCV 指数专门用于评估密度聚类质量，其计算公式如下：

$$DBCV = \frac{1}{N} \sum_{C_i \in C} [\sum_{x \in C_i} (\rho_{valid}(x) - \rho_{noise}(x))] \quad (7)$$

其中 C 是排除噪声点后所有有效簇的集合， N 代表有效簇的数量， $\rho_{valid}(x)$ 是样本 x 到最近有效簇的最小互达密度（Mutual Reachability Density）， $\rho_{noise}(x)$ 是样本 x 到噪声区域的密度阈值。DBCV 指数的取值范围也为[-1, 1]^[29]。计算本研究 DBCV 值为 0.326，基线阈值为 0.3，见图 9，说明聚类结构在统计学层面可接受。

```

=== 聚类质量评估 ===
轮廓系数: 0.529 (范围[-1,1], 越接近1越好)
DBCV指数: 0.326 (范围[-1,1], 越接近1越好)

```

图9 用 Python 计算的轮廓系数和 DBCV 指数

通过提取各组的核心代表词并分析的方式进一步总结各分组的主题内容。各组的代表性词语选择基于组内语义中心度 (Intra-cluster Semantic Centrality) 的计算, 中心度范围为 0~1, 越接近 1 则中心性越强。将各组词语的中心度排序, 并根据每组词语的数量动态确定选取的代表词语数量, 以确保每组都能有代表性的词语呈现。表 2 所示为各组总体统计情况及代表词情况, 括号中是词语中心度。可看出除噪声点外, 所有词语可分为共 27 类 (编号为 0~26)。

表 2 各组情况及代表词 (及代表词中心度)

分组 ID	词语数量	代表词及中心度
-1	140	噪声点
0	27	农村 (0.497)、时代 (0.494)、人民 (0.491)、事业 (0.491)、思想 (0.490)
1	60	产业化 (0.939)、西红柿 (0.938)、利益 (0.938)、报道 (0.937)、产业链 (0.937)
2	19	样本 (0.922)、汇报 (0.900)、课程 (0.899)
3	49	特征 (0.858)、采样 (0.849)、软件 (0.840)、西瓜 (0.837)、酶 (0.835)、鉴定 (0.832)、虫子 (0.831)
4	37	有机 (0.692)、病 (0.672)、药剂 (0.668)、防治 (0.668)、虫 (0.650)
5	28	距离 (0.802)、目标 (0.799)、计算 (0.791)、模型 (0.777)
6	86	建立 (0.779)、鼓励 (0.740)、平台 (0.739)、实施 (0.737)、节约 (0.730)、竞争力 (0.729)、体系 (0.725)、行业 (0.712)、责任 (0.710)
7	213	水量 (0.770)、植物 (0.768)、生长 (0.764)、提高产量 (0.762)、幼苗 (0.760)、透气性 (0.759)、植株 (0.757)、清水 (0.757)、珍惜 (0.756)、修复 (0.755)、伤害 (0.754)、基础设施 (0.752)、秸秆 (0.751)
8	11	细菌 (0.856)、菜 (0.843)、堆肥 (0.824)
9	21	微信 (0.786)、习惯 (0.751)、事情 (0.738)
10	23	工人 (0.857)、疼 (0.826)、着急 (0.818)
11	38	评价 (0.870)、手段 (0.869)、高产 (0.866)、差距 (0.863)、追肥 (0.861)、防控 (0.852)
12	13	睡觉 (0.708)、毕业 (0.671)、喝 (0.660)
13	12	文献 (0.721)、阅读 (0.705)、配合 (0.621)
14	21	火龙果 (0.972)、称肥 (0.966)、打水 (0.966)、出花 (0.966)
15	92	土 (0.902)、显著性 (0.895)、下雨 (0.892)、钾 (0.892)、晒 (0.891)、花 (0.890)
16	15	食物 (0.757)、催芽 (0.746)
17	21	取土 (0.898)、电视台 (0.885)、记者 (0.880)
18	22	收获 (0.874)、经验 (0.864)、采取 (0.844)、工作效率 (0.811)
19	47	知识 (0.740)、研究 (0.694)、方法 (0.689)、利用 (0.688)、总结 (0.682)、帮助 (0.678)
20	18	运行 (0.961)、测试 (0.959)、寻找 (0.958)、调试 (0.958)
21	34	实际 (0.883)、方案 (0.879)、计划 (0.877)、学到 (0.869)、交流 (0.865)
22	23	感谢 (0.893)、测 (0.871)、指标 (0.870)、海藻 (0.866)
23	10	准备 (0.733)、土壤 (0.652)、负责 (0.646)

24	45	剪辑 (0.858)、菌肥 (0.857)、宣讲 (0.849)、故事 (0.841)、录制 (0.841)
25	14	开会 (0.927)、院士 (0.893)
26	49	全球 (0.792)、讲解 (0.786)、农户 (0.773)、梳理 (0.770)、资金 (0.764)、产能 (0.762)

同时，通过词语进行聚类后得到的主题散点图，如图 10 所示。图中各个主题间的界限较为分明，展现了良好的聚类效果。

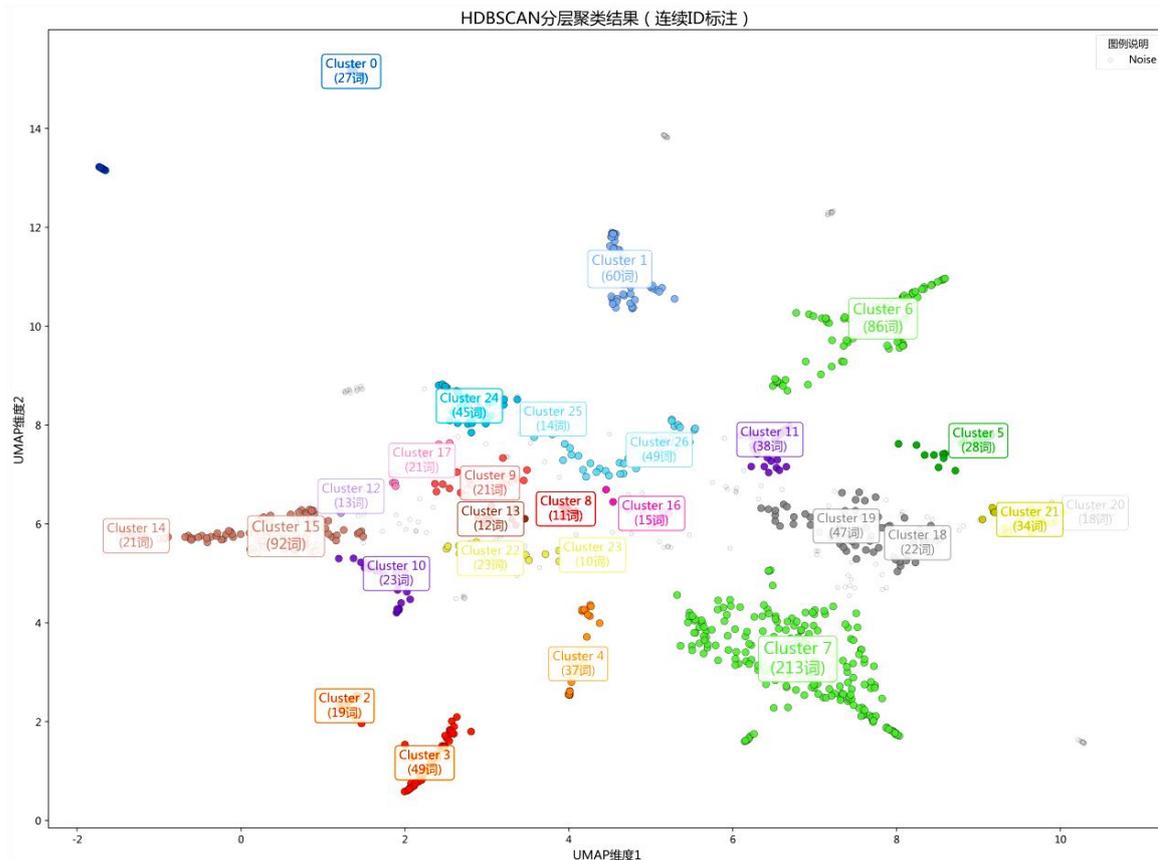


图 10 主题散点图

(1) 分析

将原聚类生成的各组词语主题编号 C，对每组词进行人工分析及去除噪声词，将每组词概括出一个主题，如表 3 所示。

表 3 各组主题及代表词

组号	主题	代表词
C0	农业政策体系与民生	农村 时代 人民 事业 思想
C1	农业市场与经济效益	产业化 西红柿 利益 报道 产业链
C2	科研数据采集与教学	样本 汇报 课程
C3	作物特征智能检测	特征 采样 软件 西瓜 酶 鉴定 虫子
C4	有机农业与病虫害防治	有机 病 药剂 防治 虫

C5	农业智能化发展	距离 目标 计算 模型
C6	管理机制与执行	建立 鼓励 平台 实施 节约 竞争力 体系 行业 责任
C7	作物栽培与土壤改良	水量 植物 生长 提高产量 幼苗 透气性 植株 清水 珍惜 修复 伤害 基础设施 秸秆
C8	农业物质与元素	细菌 菜 堆肥
C9	社交与生活	微信 习惯 事情
C10	劳动安全、应急管理与健康	工人 疼 着急
C11	农业生产力研究	评价 手段 高产 差距 追肥 防控
C12	日常生活	睡觉 毕业 喝
C13	学术资源与研究	文献 阅读 配合
C14	种植技术	火龙果 称肥 打水 出花
C15	土壤环境与元素研究	土 显著性 下雨 钾 晒 花
C16	田间作物管理	食物 催芽
C17	劳作与宣传工作	取土 电视台 记者
C18	农业生产经验与效率提升	收获 经验 采取 工作效率
C19	知识管理与应用	知识 研究 方法 利用 总结 帮助
C20	系统运行与维护	运行 测试 寻找 调试
C21	实际问题的解决方案与学习交流	实际 方案 计划 学到 交流
C22	生物指标检测技术	感谢 测 指标 海藻
C23	农业准备工作	准备 土壤 负责
C24	技术传播与媒体制作	剪辑 菌肥 宣讲 故事 录制
C25	学术会议及交流	开会 院士
C26	国际农业产能协同	全球 讲解 农户 梳理 资金 产能

(2) 归纳

考虑到 27 个原始主题（C 类）数量较多，为便于后续分析，继续将 27 类词组进行归纳。采用分阶段归纳的方式：首先将语义关联紧密的簇合并为 10 个中级主题（B 类），再进一步整合为 5 个核心主题（A 类）。归纳依据为 UMAP 降维主题散点图中簇的空间邻近性（反映主题语义关联），并通过 Python 自动化工具辅助完成合并，兼顾分析的细致性与可操作性。如表 4 所示。第二次生成的主题散点图，如图 11 所示。

表 4 第一次归纳情况及主题

组号	包含	主题	组号	包含	主题
B0	C8、C11、C16、 C23、C26	农业生产要素优化与国际 产能协作	B5	C0	农业政策与民生保障 体系
B1	C5、C18、C19、	农业智能技术集成与知识	B6	C6	农业管理机制

	C20、C21	系统构建			
B2	C14、C15	种植技术与环境调控	B7	C2、C3	农业科研数据化与智能检测
B3	C4、C7	可持续种植与生态治理	B8	C24、C25	技术传播与学术交流
B4	C9、C10、C12、C13、C17、C22	科研人员生活健康管理与基础研究	B9	C1	农业市场化与产业链

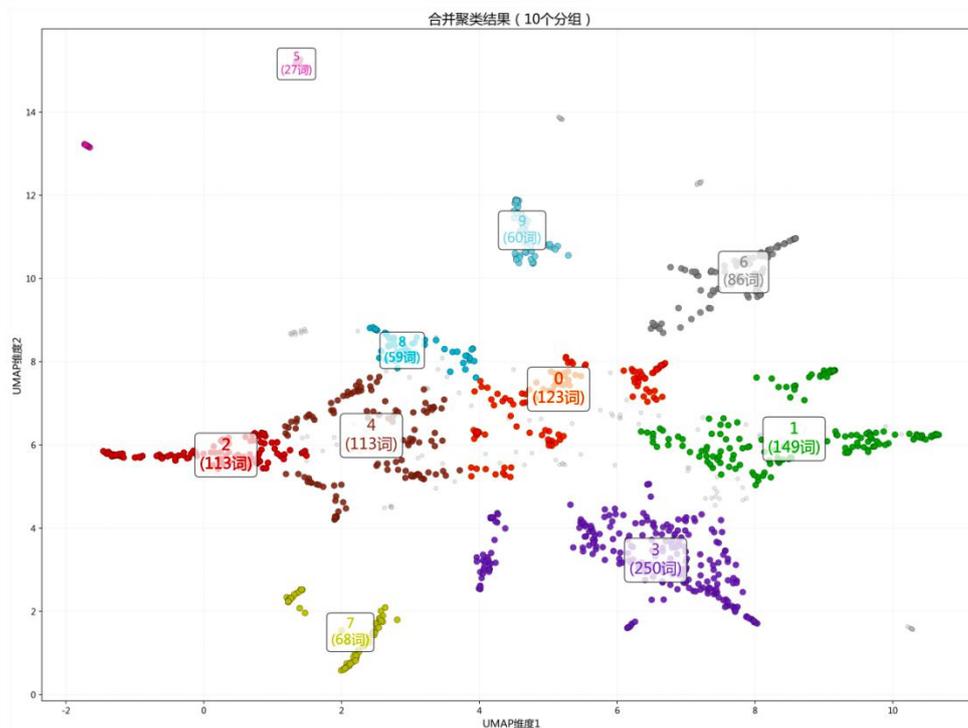


图 11 第二次生成的主题散点图

再进一步归纳，同样通过 Python 将临近的簇合为一类，如表 5 所示。第三次生成的主题散点图如图 12 所示。

表 5 第二次归纳情况及主题

组号	包含 B	主题
A0	B0、B6、B9	农业经济体系、市场与全球化
A1	B2、B4、B8	农业生产技术应用与社会化服务
A2	B5	农业政策与民生保障制度
A3	B3、B7	农业生态安全与科研体系
A4	B1	农业智能系统与知识工程

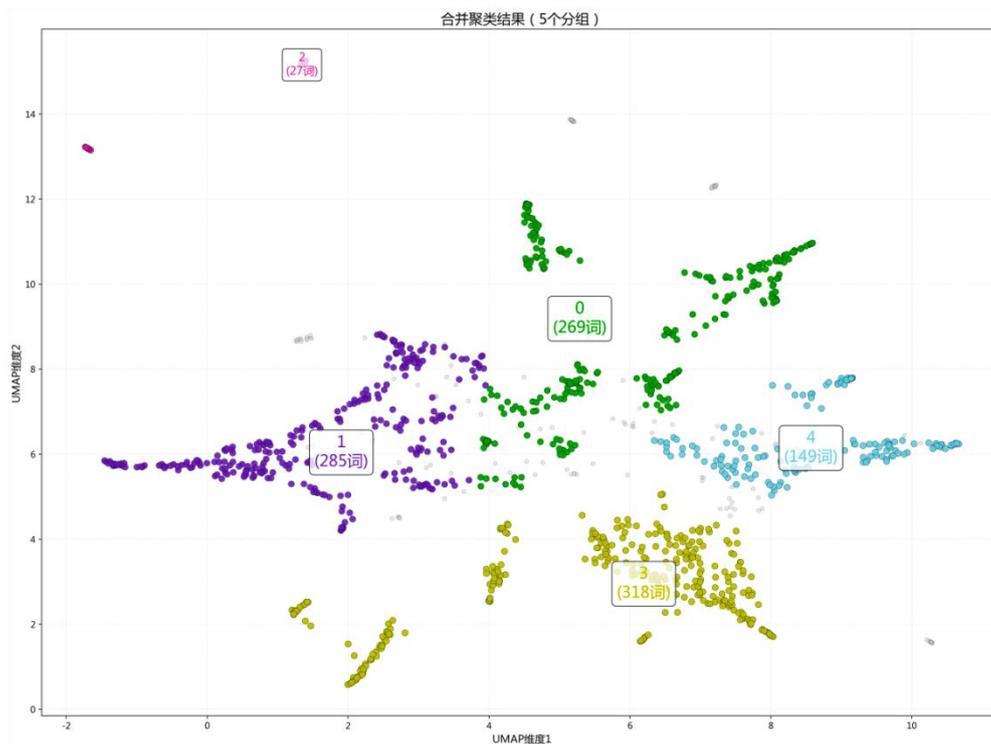


图 12 第三次生成的主题散点图

将具体的词语归类, 统计各主题词语数量及占比, 将原聚类生成的各组词语主题编号 C, 第一次归纳主题编号 B, 第二次归纳主题编号 A, 如表 6 所示。

表 6 主题数量统计表

第二次归纳	第一次归纳	原主题	占比	占比	占比
A0 农业经济体系、市场与全球化	B0 农业生产要素优化与国际产能协作	C8 农业物质与元素	1.0%	11.7%	25.7%
		C11 农业生产力研究	3.6%		
		C16 田间作物管理	1.4%		
		C23 农业准备工作	1.0%		
		C26 国际农业产能协同	4.7%		
	B6 农业管理机制	C6 管理机制与执行	8.2%	8.2%	
B9 农业市场化与产业链	C1 农业市场与经济效益	5.7%	5.7%		
A1 农业生产技术应用与社会化服务	B2 种植技术与环境调控	C14 种植技术	2.0%	10.8%	27.2%
		C15 土壤环境与元素研究	8.8%		
	B4 科研人员生活健康管理	C9 社交与生活	2.0%	10.8%	

	理与基础研究	C10 劳动安全、应急管理与健康	2.2%	5.6%	
		C12 日常生活	1.2%		
		C13 学术资源与研究	1.1%		
		C17 劳作与宣传工作	2.0%		
		C22 生物指标检测技术	2.2%		
B8 技术传播与学术交流	C24 技术传播与媒体制作	4.3%	5.6%		
	C25 学术会议及交流	1.3%			
A2 农业政策与民生保障制度	B5 农业政策与民生保障体系	C0 农业政策体系与民生	2.6%	2.6%	2.6%
A3 农业生态安全与科研体系	B3 可持续种植与生态治理	C4 有机农业与病虫害防治	3.5%	23.9%	30.3%
		C7 作物栽培与土壤改良	20.3%		
	B7 农业科研数据化与智能检测	C2 科研数据采集与教学	1.8%	6.5%	
		C3 作物特征智能检测	4.7%		
A4 农业智能系统与知识工程	B1 农业智能技术集成与知识系统构建	C5 农业智能化发展	2.7%	14.2%	14.2%
		C18 农业生产经验与效率提升	2.1%		
		C19 知识管理与应用	4.5%		
		C20 系统运行与维护	1.7%		
		C21 实际问题的解决方案与学习交流	3.2%		

由表 6 可知，在科技小院日志中有关主题 A0 农业经济体系、市场与全球化的词语数量占比为 25.7%，A1 农业生产技术应用与社会化服务占比为 27.2%，A2 农业政策与民生保障制度占比为 2.6%，A3 农业生态安全与科研体系占比为 30.3%，A4 农业智能系统与知识工程占比为 14.2%。数据表明，科技小院的实践重心呈现差异化特征：生态安全与科研体系（A3）、生产技术应用（A1）及经济体系（A0）方面需求较多，而政策制度（A2）与智能系统（A4）



图 14 图书馆数据资源服务对科技小院的保障路径

(1) 农业经济体系支撑服务（服务于 A0 主题）

① 数据资源收集与整合

建立国际农产品市场动态数据库：收集全球主要农产品的价格波动、供需变化、贸易政策等信息，形成农业物质与元素等相关主题的动态数据库，为科技小院科研人员提供最新的国际市场资讯，助力其把握全球化背景下的农业经济走向。

建立国内农业区域经济数据库：按不同农业产区、经济区域整理农业生产、销售、产业结构等数据，帮助科研人员分析国内农业经济体系的区域特点和发展趋势。

② 信息服务与咨询

定期推送市场分析报告：基于上述数据库，定期生成并推送农业经济市场分析报告，涵盖农产品进出口形势、国内农业产业竞争力等内容，满足科研人员对农业经济体系、市场与全球化相关信息的需求。

个性化咨询服务：针对科研人员在农业经济全球化研究中的个性化问题，如国际农业产

能合作项目的可行性分析等，提供专业的信息咨询和数据支持。

(2) 生产技术应用及社会化支持服务（服务于 A1 主题）

① 搭建技术资源平台

建立农业生产技术知识库：整合 B4 科研人员健康管理与基础研究，B2 种植技术与环境调控，包括种植、养殖、加工等各领域的先进技术和实用方法，形成可检索、可共享的知识库，方便科技小院科研人员随时获取。

建立农业技术交流社区：建立线上交流平台，促进科研人员之间关于农业生产技术应用交流，分享经验、讨论问题、开展合作研究。

② 技术推广与培训

线上技术培训课程：开发或收集农业生产技术相关的培训课程，如 B8 技术传播与媒体制作等，通过视频、图文等形式，对科技小院科研人员进行远程培训，提高其技术应用能力。

线下技术示范与推广：与科技小院合作，定期组织农业生产技术示范活动，将先进的生产管理技术等在地进行展示和推广，推动技术在农业生产中的应用。

(3) 政策民生保障制度配套服务（服务于 A2 主题）

① 政策法规数据库建设

农业政策法规库：全面收集国家和地方的农业政策法规文件，涵盖农业补贴、土地流转、农产品质量安全等方面，为科研人员研究 B5 农业政策民生保障体系等提供政策依据和数据支持。

政策解读与案例分析：对重要的农业政策法规进行解读，并收集相关的实施案例，帮助科研人员理解政策内涵和应用场景。

② 民生需求调研与反馈

民生需求调研平台：搭建调研平台，收集农民和农业企业等对农业政策民生保障的需求和意见，为科研人员开展相关研究提供一手数据，调研各种农业技术如何更好地服务民生保障。

建立政策反馈渠道：协助科研人员将研究成果和政策建议通过有效渠道反馈给政府部门，促进农业政策的优化和完善，保障民生需求。

构建“高校图书馆—科技小院—地方政府”协同模式，为高校资源下沉提供标准化操作模板。

(4) 生态安全科研协同服务（服务于 A3 主题）

① 生态安全数据监测与分析

建立农业生态环境监测数据库：整合土壤环境与元素研究、作物栽培与土壤改良等方面的数据，建立长期的农业生态环境监测数据库，为科研人员评估农业生态安全状况提供数据支持。

建立生态安全风险评估系统: 基于监测数据开发生态安全风险评估系统,对农业生产中的生态风险进行预警和评估,如针对农业科研数据化与智能检测等需求,提供智能化的分析工具。

② 科研体系支持与协作

建立科研管理与协作平台: 为科技小院的科研项目提供全流程管理服务,包括项目申报、进度跟踪、成果管理等,促进 A3 农业生态安全与科研体系的高效运行。

做好跨学科科研协作支持: 利用图书馆的资源和平台优势,促进农业生态安全领域的跨学科协作,如联合环境科学、生态学等学科的科研人员,共同开展相关研究。

实现学科馆员协同服务: 结合驻点指导(可基于“兴农书屋”等实地服务)与线上响应(如及时的文献传递),缩短技术转化链条。

(5) 智能知识工程赋能服务(服务于 A4 主题)

① 智能系统数据支撑

建立农业大数据中心: 收集和整合农业生产、市场、生态等多领域的的数据,为农业智能系统与知识工程构建提供海量、全面的数据基础,如农业生产经营效率与效益提升等智能系统的开发所需的数据。

数据预处理与标引服务: 对收集到的原始数据进行清洗、预处理和标注,提高数据质量,使其更适合智能系统的训练和应用,如为农业智能技术集成与知识系统构建提供高质量的数据资源。

② 知识工程服务与推广

构建农业知识图谱: 基于农业领域的知识体系,构建农业知识图谱,将分散的知识进行关联和整合,为科研人员提供更全面、系统的知识服务,如在系统运行与维护等方面的知识支持。

智能系统推广培训: 协助科技小院推广和应用农业智能系统,通过培训和技术支持,提高科研人员和农业从业者对智能系统的使用能力,如实际问题的解决方案与学习交流等智能系统的推广应用。

上述分析表明,图书馆通过 A0~A4 主题维度的差异化服务策略,能够有效响应科技小院在农业经济、生产技术、政策民生等领域的需求。然而,单一主题维度的服务供给可能存在一些效能瓶颈:一是需求关联断裂(如 A3 生态数据与 A1 技术应用间的知识孤岛);二是资源复用率低下(如 A4 智能系统可能无法充分调用 A0 农产品市场的数据);三是服务响应迟滞(如 A2 政策服务可能难以动态适配 A0 市场合作需求)。为完善服务体系,接下来进一步分析各服务主体的协同作用。

(6) 各保障方式间的协同作用机制

各服务主题(一)~(五)并非孤立存在,而是相互关联、相互促进,共同形成一个有机整体。以下是具体剖析多主题保障方式间复杂的非线性交互机制,以及每一项保障服务主题与其他主题的作用机制(见表 7)。

表 7 图书馆数据资源各保障方式间的协同作用机制

	A0 农业经济体系支撑服务	A1 生产技术社会化应用支持服务	A2 政策民生保障制度配套服务	A3 生态安全科研协同服务	A4 智能知识工程赋能服务
A0		与 A0 服务协同, 将国际市场动态和区域经济分析结果融入技术推广策略中, 提高技术推广的针对性和有效性。	与 A0 服务协同, 确保国际化决策符合国内政策导向和民生需求。	与 A0 服务协同, 将生态安全形势纳入全球经济和区域经济分析中, 提升决策的全面性和科学性。	与 A0 服务协同, 利用大数据和认知计算技术提升国际市场动态和区域经济分析的准确性和效率。
A1	为 A1 服务提供农业国内外市场趋势和区域产能需求信息, 指导技术推广的方向和重点		为 A1 服务提供政策指导和支持, 促进技术的合规应用和推广。	为 A1 服务提供生态治理技术和案例, 促进生产技术的生态化转型。	为 A1 服务提供智能化的技术推广策略和建议, 提高技术推广的精准度和效果。
A2	为 A2 服务提供国际农产品政策动态和区域经济分析结果, 作为政策制定和优化的参考。	为 A2 服务提供技术应用案例和反馈, 作为政策制定和调整的依据。		与 A2 服务协同, 推动生态治理政策的科研支撑和制度优化。	与 A2 服务协同, 利用智能技术优化政策制定和实施过程, 提升政策的有效性和针对性。
A3	为 A3 服务提供全球生态安全形势和区域经济影响评估, 助力生态监测数据库的开发和跨学科协作平台的构建。	与 A3 服务协同, 将田间管理案例和生态治理技术相结合, 推动生态安全科研的协同发展。	与 A3 服务协同, 推动生态治理政策的制定和实施, 保障生态安全科研的顺利进行。		为 A3 服务提供智能化的生态监测和风险评估工具, 支持生态安全科研的深入发展。
A4	为 A4 服务提供全球经济数据和农业大数据的整合需求, 促进农业大数据中心和知识图谱系统的建设。	为 A4 服务提供丰富的技术数据和知识资源, 支持农业大数据中心和知识图谱系统的不断完善。	为 A4 服务提供政策数据和分析结果, 支持智能知识工程在符合政策框架下的赋能服务。	为 A4 服务提供生态监测数据和跨学科研究成果, 丰富农业大数据中心和知识图谱系统的内容。	

① 农业经济体系支撑服务（A0）与其他服务的协同

农业经济体系支撑服务作为服务体系的起点, 通过构建国际农产品市场动态数据库与区域经济分析系统, 为生产技术社会化应用支持服务（A1）、政策民生保障制度配套服务（A2）、生态安全科研协同服务（A3）以及智能知识工程赋能服务（A4）提供宏观的经济背景和市场动态信息。

② 生产技术社会化应用支持服务（A1）的协同作用

生产技术社会化应用支持服务通过搭建技术知识库和产学研协同平台，促进农业技术的推广和应用，还与其他服务产生紧密的协同作用。

③ 政策民生保障制度配套服务（A2）的协同影响

政策民生保障制度配套服务通过建立政策法规库和民生需求反馈机制，为其他服务提供政策支持和制度保障。

④ 生态安全科研协同服务（A3）的跨领域融合

生态安全科研协同服务通过开发生态监测数据库和跨学科协作平台，实现生态安全与农业科研的深度融合。

⑤ 智能知识工程赋能服务（A4）的全面支撑

智能知识工程赋能服务通过建设农业大数据中心和知识图谱系统，为其他服务提供智能化的数据支持和知识服务。

综上所述，各子服务体系之间可以通过信息共享、资源互补、技术融合等方式形成紧密的协同作用机制。这种协同作用不仅提升了整个服务体系的整体效能和服务质量，还可以为科技小院提供全面、高效、智能的数据资源服务保障。

3.5.2 保障路径的优化

党的二十大报告指出“教育、科技、人才是全面建设社会主义现代化国家的基础性、战略性支撑。”高校尤其是“双一流”建设大学必须勇担政治责任和历史使命，作为科技第一生产力、人才第一资源、创新第一动力的重要结合点，必须主动参与产学研深度融合中，为有效衔接、贯通实施科教兴国战略、人才强国战略、创新驱动发展战略，实现教育、科技、人才一体推进良性循环探索有效路径^[30]。当前，我国正积极采用“政产学研用”融通共建模式以促进科技小院的快速发展。在此模式下，农林院校及其图书馆作为“学”的重要组成部分，与其余环节紧密相连、共同发展。同时，农林院校及其图书馆应充分发挥自身优势，为科技小院在人才培养、科学研究、社会服务等方面提供有力支持，进而推进科技小院服务三农、人才培养和场域效应的一体化。

一个科技小院就犹如一个“场”，理论的验证、科技的创新、技术的推广、人才的培养等，这些事物在交流、互动、冲突中相互作用、相互促进，并衍生出更大的“场域效应”，辐射带动更多人、更广领域，为推动农业农村现代化不断积聚新动能。发挥好科技小院的“场域效应”，使之成为推进乡村振兴的有力引擎，至关重要^[31]。科技小院的工作模式为“农业科研+科技服务+人才培养”，这一点决定了科技小院也是储备人才的重要路径，因此在依托科技小院自身优势发挥“场域”效应，因地制宜扩大服务三农的功能的同时，也应注重科研人员能力的提升。对应地，科技小院工作的全面提升，利于致富一方百姓、繁荣一方乡村，对于“政产学研用”也有一定促进作用。

当前图书馆数据资源服务中的“数据”不仅是基础性、战略性的资源，更成为变革传统方法和工具的关键推动力。数据资源的应用不仅可以通过交易、赋能等方式创造新的商业模式和增值机会，还可以引领组织转型、提升效率 and 创新能力。在这一背景下，图书馆资源服务保障的重要性日益凸显。通过将线上线下各类资源整合转化为同等重要的资产，高校图书馆能够为广大用户提供丰富多样的信息服务，满足不断增长的学术和研究需求。科技小院作

为其中的重要用户群体，也能从中获益。具体机理如图 15 所示。

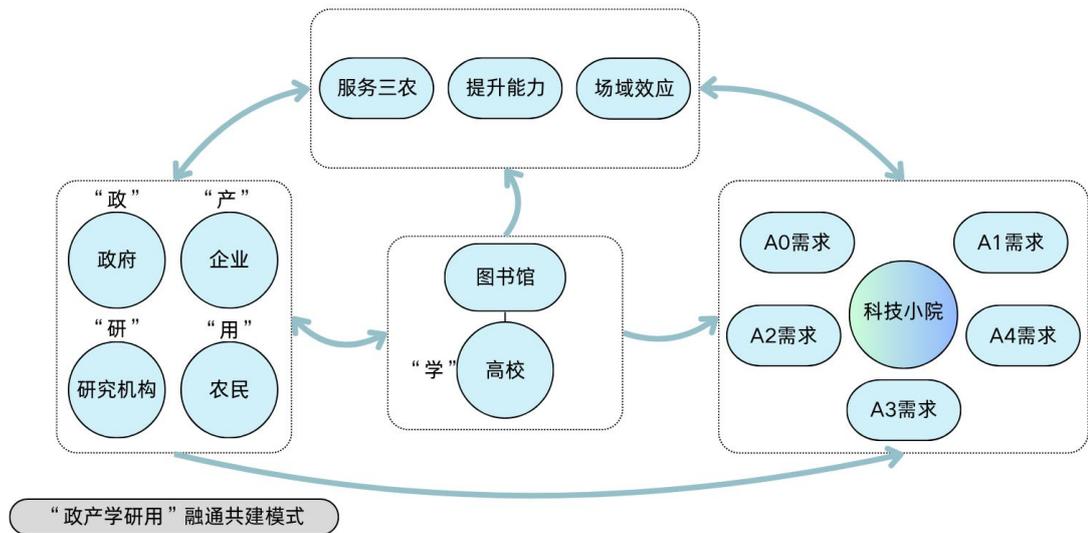


图 15 图书馆资源服务保障影响机理模型

从更深远的视角来看，农业林高院图书馆不仅要深化服务内涵，更在于拓宽服务边界，确保图书馆能够持续适应并引领知识服务的新趋势，为科技小院研究人员创造良好的知识获取与创新环境。结合构建的图书馆数据资源服务保障路径，增强高校资源服务保障能力应聚焦于以下四个方向。

(1) 需求导向的服务模式创新

需求导向的服务模式具体做法即需求驱动及协同发展的图书馆数据资源服务具体保障路径。在当前科技小院“政产学研用”融通共建的模式下，图书馆需结合其含有“研究生—农民—企业—政府”的多元主体，通过统计分析得出的科研人员的需求，引申到农民、企业、政府对于图书馆在科技小院建设与运行方面的期待，重构服务模式，形成具有创新性的需求导向的数据资源服务保障路径。

保障方式应加以扩展和延伸，图书馆可建立对应各主体的分层响应机制。首先对于政府和企业，可通过提供农业相关数据库与政策解读服务，提供产业布局建议与政策风险评估；第二对于科研团队，可通过整合生态数据与智能分析工具，支持科研人员跨学科科研攻关；也可基于大数据分析预判区域农业趋势，提前配置生态防治方案库与应急技术指南。第三是对于当地农民，将技术知识库内容转化为通俗易懂的多媒体教程，降低技术采纳门槛。

除此之外，图书馆数据资源保障服务要明确功能和定位，要动态地了解“研究生—农民—企业—政府”这些多元主体的需求类型和内容，真正做到为科技小院主体服务目标。同时，要意识到资源价值属于显性影响因素。在高校图书馆资源服务保障中，资源是决定服务能够展开的基础，高校图书馆要充分利用自身馆藏资源优势，深挖馆藏资源创造性地转化为符合科技小院需要的教学资源、娱乐资源等，形成可供科技小院的多主体全方位的知识服务的特色主题资源。不仅提升了资源的价值，同时盘活馆藏资源，使之丰富充实，提升高校图书馆馆藏资源的利用效能。

(2) 文化传承与科技服务相融合

科技小院作为乡村振兴文化载体的一部分,其服务保障需突破传统技术支持的单一维度,构建“技术赋能+文化增值”的双轨模式。

在高校图书馆资源服务保障中,高校图书馆扮演着提升用户文化自信、推动社会文化发展的重要角色。挖掘科技小院主体的意愿是服务的动力,而凝练科技小院所处地区的当地文化特色和文化价值,则是高校图书馆资源服务的方式、方法和手段。高校图书馆可以通过挖掘科技小院所在地域的非物质文化遗产(如农谚、传统耕作技艺),通过知识图谱技术构建技术与文化关联数据库,并根据科技小院所在省市地区的历史、文化和自然景观等特点,不断开拓数据资源,尤其是将当地特有和创新的资源以及异质资源(如视频资源等)有机融合,建设特色馆藏资源^[32],盘活当地文化资源,提升农民文化认同感;也可以依托图书馆的文献组织能力,协助科技小院策划“乡村书屋+技术驿站”融合空间,既提供种植技术手册,也配置农耕文化读本,丰富农民的精神文化生活,推进乡村振兴的精神文明共建服务,也是对高校图书馆资源的拓展和创新,为地方文化的传承和发展贡献力量。

(3) 制度规范与长效发展保障

目前由于缺乏服务制度和流程,也缺少服务反馈机制,图书馆资源服务存在碎片化的难题,为解决这一问题,图书馆可以通过制度创新构建可持续保障体系。不仅可以加强资源建设和服务保障,让农民成为受益的主体,而且还可以强化组织创新,以党建为引领更好地发挥科技助农作用,扩展服务内容。

首先可建设标准化服务模式,明确各种服务的响应时限、质量标准和协作规则。建立一套确保数据质量、安全和隐私保护的规范和流程,定期进行资源服务和保障的评价与持续改进服务策略和保障资源质量以及技术手段的迭代更新,才能持久为科技小院人才培养提供有力扶持。第二,图书馆可以依托未来学习中心,与当地图书馆、研究机构和当地政府以及企业良好建立合作关系共享资源和服务。既能丰富图书馆数据资源,也能使本校园学生获得更多资源。第三,建立动态评估与反馈机制,如建立由图书馆、科技小院、地方政府组成的联合评估小组,定期审查政策服务落地效果、技术传播覆盖率等核心指标。第四,可以拓展区域服务网络,以重点科技小院为枢纽节点,构建区域图书馆资源协作联盟。

(4) 场域效应下的服务价值提升

科技小院“场域效应”的本质上是资源、知识、行动者关系的动态重构。为使图书馆服务价值在此效应下得到提升,不再仅作为资源供给的客体,图书馆可以通过构建农业知识工程体系,将碎片化的技术规程、生态监测数据与政策规范进行知识重组,让图书馆成为乡村振兴知识生态的共建者;也可以依托知识的数据整合与调度能力,将各种主题服务模块嵌入地方农业数字化转型进程,如通过将智能灌溉技术指南等主题服务嵌入县域智慧农业平台,构建“技术供给—政策适配—市场对接”的全链条服务网络;还可以通过提炼科技小院实践案例,构建农业现代化伦理知识库,为乡村振兴战略提供兼具技术合理性与社会正当性的价值范本。

4. 结论与展望

4.1 研究结论

4.1.1 揭示需求特征

基于科技小院实践日志的文本挖掘与主题分析，研究揭示了 A0~A4 五维需求主题的分布特征，需求占比较高的为：农业生态安全与科研体系（A3，30.3%）、生产技术应用（A1，27.2%）、经济全球化协作（A0，25.7%）。而政策制度（A2，2.6%）与智能系统（A4，14.2%）的需求显性化程度较低。这一分布表明，科技小院当前实践重心聚焦于生态治理、技术推广及国际产能协作等应用领域，而政策适配与智能技术融合仍存在需求表达不足的问题。

4.2.2 构建保障路径

针对需求层级特征，提出图书馆数据资源服务的五维保障路径：农业经济体系支撑服务（服务于 A0）、生产技术社会化应用支持服务（服务于 A1）、政策民生保障制度配套服务（服务于 A2）、生态安全科研协同服务（服务于 A3）、智能知识工程赋能服务（服务于 A4）。并列各路径间通过数据共享、资源互补与技术融合可形成的协同作用机制。

4.2.3 优化保障路径

为增强服务可持续性，结合科技小院“政产学研用”融通共建模式和科技小院场域效应的特点，提出四重优化方向：

一是需求导向的服务模式创新，即通过具体对接政府、科研团队与农民需求，建立对应各主体服务的分层响应机制。

二是文化遗产与科技服务融合，即通过构建技术与文化关联数据库，推动乡村文化资源活化与技术传播协同。

三是制度规范与长效发展保障，即制定标准化服务流程与动态评估机制，构建区域图书馆资源协作联盟。

四是场域效应下的服务价值提升，即构建农业知识体系、依托资源整合能力嵌入地方数字化转型进程、提炼科技小院实践案例，构建农业现代化伦理知识库等。

4.2 研究展望

研究存在以下局限和待深化方向：

一是样本数据方面。由于收集渠道单一及网站日志上传不足，当前样本局限于单一高校的 28 个科技小院的 1906 篇日志，且日志分布不均匀。未来可纳入多区域、多类型机构的日志数据，建立跨层级、跨类型的科技小院日志数据库进行文本分析，增强需求分析的普适性和代表性。

二是研究方法方面。在聚类后采用了人工主题识别的方式，有一定的主观性，未来可优化主题聚类算法，探索深度学习模型在需求语义解析中的应用，提升主题识别的细粒度。并通过动态更新主题库，为服务匹配提供更加具体的决策支持。

三是实践验证方面。研究形成的保障路径停留在理论层面，未来可以推动保障路径的试点应用，通过动态追踪评估服务效能，提炼可量化的服务标准，形成可复制的标准化服务范式，助力科技小院模式在国内外规模化推广。

5. 参考文献

- [1] 李乾, 张福锁, 焦小强, 等. 科技小院创新发展: 现实需要、功能定位与运行保障机制[J]. 农业现代化研究, 2023, 44(01): 1-9.
- [2] 高云, 张卓文, 赵星雨, 等. 科技小院育人模式赋能高校“大思政课”建设的路径探析[J]. 智慧农业导刊, 2025, 5(08): 146-149. DOI: 10.20028/j.zhnydk.2025.08.034.
- [3] 中中国农业大学委员会. 解民生、治学问、育英才的科技小院[EB/OL]. 求是网, [2024-04-16].
- [4] 陈思琪, 张梦园. 乡村振兴视域下科技小院赋能农业现代化对策研究[J]. 农业经济, 2025(04): 117-119.
- [5] 宋广蕊, 曹哲. 科技小院赋能农业农村现代化的中国经验与世界价值[J]. 西南金融, 2025(02): 56-66.
- [6] 中华人民共和国国务院. 中共中央国务院关于学习运用“千村示范、万村整治”工程经验有力有效推进乡村全面振兴的意见(中央一号文件)[R]. 2024.
- [7] 李明. 为强农兴农贡献青春力量——深入学习习近平总书记给中国农业大学科技小院的同学们的回信[J]. 中国农业大学学报(社会科学版), 2024, 41(03): 16-25. DOI: 10.13240/j.cnki.caujsse.2024.03.001.
- [8] 章成志, 储新龙. 中国人文社科领域研究方法应用的实证研究——基于大规模博士学位论文的调查[J]. 情报理论与实践, 2024, 47(05): 48-57. DOI: 10.16353/j.cnki.1000-7490.2024.05.006.
- [9] 陈晓云. 文本挖掘若干关键技术研究[D]. 上海: 复旦大学, 2005.
- [10] 李冬. 聚类分析的快速算法和确定类数的研究[D]. 西安: 西安电子科技大学, 2024. DOI: 10.27389/d.cnki.gxadu.2024.000197.
- [11] He-B-Ounis-I Lo T W. Automatically Building a Stopword List for an Information Retrieval System[J]. Journal of Digital Information Management, 2005, 3(1): 3-8.
- [12] 江兆中. 基于语境和停用词驱动的中文自动分词研究[D]. 合肥: 合肥工业大学, 2010.
- [13] 官琴, 邓三鸿, 王昊. 中文文本聚类常用停用词表对比研究[J]. 数据分析与知识发现, 2017, 1(03): 72-80.
- [14] 王艳, 王胡燕, 余本功. 基于多特征融合的中文文本分类研究[J]. 数据分析与知识发现, 2021, 5(10): 1-14.
- [15] 王勋涛. 几种高频词阈值计算方法的实证研究——以国内高校信息素养教育研究数据为例[J]. 晋图学刊, 2023(03): 30-38.
- [16] 宋坤莹, 冯琳琳, 王峥, 等. 词频和语境多样性在中文词汇识别中的相对重要性: 基于 ERP 的实验证据[J]. 心理与行为研究, 2024, 22(04): 433-441.
- [17] Salton G, Buckley C. Term-weighting approaches in automatic text retrieval[J]. Information Processing & Management, 1988, 24(5): 513-523.
- [18] 陈挺, 李国鹏, 王小梅. 基于 t-SNE 降维的科学基金资助项目可视化方法研究[J]. 数据分析与知识发现, 2018, 2(08): 1-9.
- [19] 任泽裕, 王振超, 柯尊旺, 等. 多模态数据融合综述[J]. 计算机工程与应用, 2021, 57(18): 49-64.
- [20] 张爱华, 孙嘉鸿. 基于 TF-IDF 算法的运营商客户投诉原因研究[J]. 北京邮电大学学报(社会科学版), 2024, 26(02): 39-49.
- [21] 武永亮, 赵书良, 李长镜, 等. 基于 TF-IDF 和余弦相似度的文本分类方法[J]. 中文信息学报, 2017, 31(05): 138-145.
- [22] 刘奕杉, 王玉琳, 李明鑫. 词频分析法中高频词阈值界定方法适用性的实证分析[J]. 数字图书馆论坛, 2017(09): 42-49.
- [23] Fabian Pedregosa, Varoquaux Gaeel, Gramfort Alexandre, et al. Scikit-learn : Machine Learning in Python[J]. JOURNAL OF MACHINE LEARNING RESEARCH, 2011, 12: 2825-2830.
- [24] 马素琴. 基于相似度的文本聚类算法研究及应用[D]. 镇江: 江苏大学, 2010.

- [25]胡吉明,郑翔. 基于主题聚类的新媒体政务互动内容摘要生成研究[J]. 数据分析与知识发现, 2022, 6(06): 95-104.
- [26]吴维芳,高宝俊,杨海霞,等. 评论文本对酒店满意度的影响:基于情感分析的方法[J]. 数据分析与知识发现, 2017, 1(03): 62-71.
- [27]张靖,段富. 优化初始聚类中心的改进 k-means 算法[J]. 计算机工程与设计, 2013, 34(05): 1691-1694.
- [28]朱连江,马炳先,赵学泉. 基于轮廓系数的聚类有效性分析[J]. 计算机应用, 2010, 30(S2): 139-141.
- [29]施格格. 优化密度峰值聚类算法及其在城市热点检测中的应用研究[D]. 芜湖:安徽师范大学, 2024.
- [30]卢建军. 加快推动产学研深度融合, 实现教育、科技、人才一体推进良性循环[J]. 中国高等教育, 2023 (11): 12-15.
- [31]吉时平. 发挥好科技小院的“场域效应” .[N]. 吉林日报,2023-06-25(1)
- [32]赵忠尧,丁燕. 图书馆文献资源建设路径创新与保障[J]. 河南图书馆学刊, 2022, 42(01): 112-114.

6. 附录：停用词表

的,了,在,是,我,有,和,不,人,都,很,你,们,他,她,它,我们,他们,这个,那个,说,看,到,上,下,去,来,吧,啊,呀,吗,呢,啊,啦,嘛,呗,也,就,还,得,着,了,过,没,无,而,让,等,但,会,再,更,挺,好,太,都,又,最,个,从,因,为,将,把,才,最,都,没,有,很,又,和,还,不,大,但,给,对,以,用,需要,出,处,进,行,着,点,些,已经,那些,一些,这里,那里,啊,吗,吧,呢,啦,呀,哦,唉,啦,呀,哟,哇,哈,唉,嘿,哦,嗯,唉呀,哎哟,哎呦,嗨,唉嘿,咯,咯瞪,呵呵,嘿嘿,哈哈,呵呵哒,嘻嘻,噗嗤,嘎嘎,嗯哼,嗯哼嗯哼,嗯哼嗯,嘶,唉呀妈呀,我的天,我的妈呀,哦哟,啊哈,嗯哼,额,呃,哎呀,哦了,今天,与,没有,地,可以,要,昨天,之前,进行,主要,以及,比较,下午,之后,多,同时,一个,方式,比较,能够,这些,影响,问题,很多,知道,不同,所以,完成,及,大家,这,时候,做,工作中,使用,还是,自己,或,非常,就是,选择,现在,通过,然后,对于,重要,一下,时,因为,晚上,相关,重要,中,开始,过程,明天,结果,但是,后,真的,看到,更加,先,继续,确保,可能,准备,发现,情况,打,包括,事情,投,完,导致,不过,前,被,具有,想,最后,其,并,能,较,减少,根据,而且,上午,最近,虽然,地理,感觉,拿,时间,小,月,高,里,可,吃,一百,地,之后,不同,开始,主要,一下,时候,工作,由于,一起,这样,两天,早上,显著,起来,有效,增加,还有,不是,方面,首先,几天,不能,那

7. 研究成果

- (1) 论文《基于扎根理论的科技小院资源保障路径分析》已发《智慧农业导刊》2025年第一期

智慧农业导刊

半月刊 2021年创刊
第5卷 第1期 总第95期 2025年1月

目次

【智慧三农】

农户数字技能评价指标体系构建与实证研究 ——以四川省为例	张婷然, 蒋雯茜, 曹心雨, 等(1)
新发展阶段我国智慧农业发展的困境与出路	仇振序(6)
大数据算法对智慧农业发展的推进研究 ——基于湖南省相关数据	胡 帆, 尹爱健, 罗 炫, 等(11)
基于葡萄种植图像识别技术的智慧农业应用	杨振强(18)
乡村振兴背景下民族地区数字乡村建设路径探究 ——以广西壮族自治区玉林市为例	周铭翔(22)
智慧农业背景下农产品出口贸易问题与对策	李一鸣, 石 楠(26)
面向智慧农业的大学生创新创业与乡村振兴耦合路径研究	甘丽群(30)

【研究视界】

基于扎根理论的科技小院资源保障路径分析	田 佳, 王 珂, 贺 玲, 等(34)
涉农高职院校推广“引鸟入林”学生高质量就业的路径研究	李伟鸿, 李 薇, 曾建英, 等(40)
涉农职业本科学生“三农”情怀培育现状及培育路径研究	于 库, 赵发娟, 屈曲波(45)
农村社会工作赋能农村文化保护路径探索	彭玉叶, 杨 章, 范文芳, 等(50)
推进乡村全面振兴视域下农民主体意识培育定位及路径研究	李明哲, 孙小利(55)
食材来源再利用调查研究分析	杨玉娜, 侯晓梅, 潘 宝, 等(59)
农业某领域研究进展与热点分析	石洪震(63)
乡村振兴目标下农村社会工作研究 ——以茶叶生产与加工技术专业为例	赵一伟(67)
“双高”建设背景下高职院校涉农专业学生创新创业能力提升路径研究	陈恩圆, 方 皓(71)

【乡村治理】

基于文化空间视角的沙河市传统村落活化利用实践及对策研究	李妙然, 王健民, 袁翠芳(75)
乡村振兴视域下推进乡村治理现代化的实践路径研究	曾 庆(79)

期刊基本参数: CN23-1613S · 2021 · 5 · A4 · 196 · 36 · P · 330.00 · 2000 · 45 · 2025-1

基于扎根理论的科技小院资源保障路径分析

田 佳, 王 珂, 贺 玲*, 杨宇红, 陈庆生
(中国农业大学 图书馆, 北京 100083)

摘 要: 新时代乡村振兴是农民生活幸福的重要组成部分, 科技小院是连接高校和田间地头, 强调多学科交叉培养和专业知识技能的关键知识, 改善农村治理资源服务网络重要载体。该文以中国农业大学科技小院的工作日志、网络调查等文本为切入点, 采用扎根理论分析方法, 从原始文本中提炼出初始概念, 然后进行开放性编码和主范畴编码, 找到概念间不同的关联, 分析各要素之间的因果、结构等内在关系, 形成理论核心类属模型, 通过分析, 提出图书馆资源服务保障助力乡村振兴的产学研协同发展、资源创新融合共享、文化提升和组织制度规范等保障路径和措施。

关键词: 高校图书馆; 资源保障; 科技小院; 扎根理论; 乡村振兴
中图分类号: G644 文献标志码: A 文章编号: 2096-9902(2025)01-0003-07

Abstract: Rural revitalization in the new era is an important part of making national rejuvenation. The sci-tech courtyard is closely connected to universities and fields, emphasizing multidisciplinary cross-cultivation and the practical expansion of professional knowledge and skills. Therefore, the guarantee of university library resources and services is particularly important. This paper takes the work diary, questionnaire and other texts of the Science and Technology College of China Agricultural University as the starting point, and uses the Grounded Theory analysis method to extract general experience from the original text, and then performs open coding and update coding to find the correlation between data elements, analyze the intrinsic relationships such as cause and effect, structure, etc. between each element, and form a theoretical core category model. Through analysis, this paper proposes guarantee paths and measures such as the industry-university-research coordinated development, integration and sharing of resource innovation, cultural improvement, and organizational innovation system seems to ensure library resource services to help rural revitalization.

Keywords: university library; resource guarantee; sci-tech courtyard; Grounded Theory; rural revitalization

2009年中国农业大学在河北曲周县白庙乡建立了一个农家小院, 第一个科技小院由此诞生。截至2023年底, 中国农业大学已在全国24个省市的91个县区牵头建立了139个科技小院, 累计引领全国建立1 178家科技小院, 2 610名专家、4 994名研究生参与了项目研究。更可喜的是, 中非科技小院群已为10多个非洲国家培养了72名农学类研究生, 并在马拉维共和国建成了4个科技小院^①。科技小院已走出国门, 与全世界分享中国的乡村振兴经验。但是, 科技小院的建设和发展仍然存在一些问题, 例如小院建设布局较散, 缺乏统筹规划; 技术服务指导多聚焦单项农业技术, 综合性全面性不足等。

强农兴农在提升科技小院运行绩效上发挥着重要作用^②。小院结出的“大成果”, 离不开科技小院师生们的艰苦奋斗, 离不开学校和社会方面的大力支持

和保障, 更离不开资源服务保障。2023年6月中国农业大学图书馆成立了“中国农业大学科技小院服务站”, 深入了解科技小院的资源需求, 主动提供各项服务, 为找到高校图书馆为科技小院提供高质量资源之路, 优化资源配置。因此, 本文旨在探究资源服务保障运行的机理, 助力科技小院的发展, 为推动乡村振兴贡献一份力量。

1 研究方法
扎根理论适用于研究以前鲜有或没有形成一致理论的问题, 是通过经验资料的不断比较、总结和提炼来形成系统理论^③。1967年, Glaser和 Strauss提出了扎根理论^④, 其特点是可以追溯到原始资料, 并以经验事实为基础。

考虑到现有文献较少, 研究采用扎根理论分析方法的艰苦努力, 离不开学校和社会方面的大力支持

基金项目: DRAA高校图书馆数字化转型背景下“智慧图书馆”建设课题(C024010A020)
第一作者简介: 田佳(2001-), 女, 硕士研究生, 研究方向为信息资源建设。
*通信作者: 贺玲(1971-), 女, 硕士, 副研究馆员, 研究方向为信息资源建设。