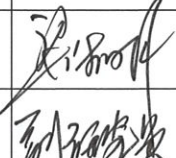
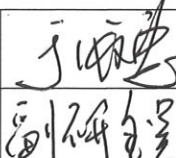
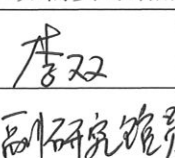


项目结题验收单

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	基于论文和专利数据的研究前沿识别方法研究——以蚕桑产业为例		
主持人	丁恩俊	职务/职称	研究馆员
所在单位	西南大学图书馆（加盖单位公章）		
专 家 意 见	<p style="text-align: center;">该项目立足于论文、专利、地理标志等多源数据，以蚕桑产业为例，进行了研究前沿识别研究，具有重要意义：一是系统性调研研究前沿概念内涵、研究前沿识别方法，构建基于多源数据的研究前沿识别方法，并以蚕桑产业为例进行实证研究；二是基于构建的研究前沿识别方法，识别出蚕桑产业的研究热点，有助于蚕桑领域研究者把握研究动向，为下一步的科学研究与技术开发、转移转化提供一定的借鉴；三是通过对研究前沿的识别、遴选、解读，以数据和可视化图表的形式把研究前沿领域展示给科研者、管理者 and 决策者，为其开展科学研究、制订科学技术发展的路线、战略和政策提供依据；四是完善图书馆的学科服务体系，为开展研究前沿分析服务、知识产权信息服务和学科发展分析服务等提供参考。</p> <p style="text-align: right;">（如需要可增加页数）</p>		
专家签字			 
职务/职称	副研究员		副研究员 副研究馆员



项目编号：2022061

CALIS 全国农学文献信息中心研究项目 结题报告

项目名称：基于论文和专利数据的研究前沿识别方法研究——以蚕桑产业为例

项目关键词：研究前沿识别 论文数据 专利数据 蚕桑产业

项目单位(盖章)：西南大学图书馆

通信地址：(详细地址含邮编)重庆市北碚区天生路 2 号 西南大学中心图书馆（邮编：400715）

项目主持人：丁恩俊

联系电话：13667696417

电子邮件：67865138@qq.com

提交日期：2023 年 5 月 20 日

(结题报告含有以下 5 部分内容，其他内容根据项目情况可增加，字数不少于 4000 字)

题目：基于论文和专利数据的研究前沿识别方法研究——以蚕桑产业为例

关键词：研究前沿识别 论文数据 专利数据 蚕桑产业

1 研究背景、目的及意义

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。必须坚持科技是第一生产力、人才是第一资源、创新是第一动力，深入实施科教兴国战略、人才强国战略、创新驱动发展战略，开辟发展新领域新赛道，不断塑造发展新动能新优势。”“加快实施创新驱动发展战略，加快实现高水平科技自立自强，以国家战略需求为导向，集聚力量进行原创性引领性科技攻关，坚决打赢关键核心技术攻坚战，加快实施一批具有战略性全局性前瞻性的国家重大科技项目，增强自主创新能力。”国务院印发的《统筹推进世界一流大学和一流学科建设总体方案》提出“高校要提高基础研究水平，争做国际学术前沿并行者乃至领跑者。推动加强战略性、全局性、前瞻性问题研究，着力提升解决重大问题能力和原始创新能力。”习近平总书记在主持召开科学家座谈会时提出的“四个面向”，其中第一个就是面向世界科技前沿。《中

华人民共和国国民经济和社会发展第十四个五年规划和2035年远景目标纲要》中也频频提到“前沿”，如“科技前沿”“前沿领域”“前沿技术”等。可以看出，研究前沿在完善国家创新体系，加快建设科技强国中具有重要的战略意义。而准确地识别研究前沿可以帮助科研者、管理者和决策者在宏观上了解科研动态，洞察科研动向，把握最新研究进展，跟踪新兴专业技术领域，从而精准地开展科研工作，将有限的精力和资源用在最恰当的地方，支持和推进科学进步。

当前开放数字环境下，数据库比较健全，论文、专利等数据都有特定的数据库提供下载，容易获取；同时，在科学研究意义上，论文、专利作为科研产出的主要形式，前者反映着基础研究，后者反映着应用研究，分析论文、专利数据可以从侧面反映出基础研究、应用研究层面上科学技术发展的前沿趋势。将论文和专利相结合识别研究前沿，是研究前沿识别上的新视角。本文立足于论文与专利数据，进行论文和专利相结合的研究前沿识别方法研究，并以蚕桑产业为例，对所构建的研究前沿识别方法进行实证研究。

本研究的意义在于：一是系统性调研研究前沿概念内涵、研究前沿识别方法，构建基于论文和专利数据的研究前沿识别方法，并以蚕桑产业为例进行实证研究；二是基于构建的研究前沿识别方法，识别出蚕桑产业的研究热点，有助于蚕桑领域研究者把握研究动向，为下一步的科学研究与技术开发、转移转化提供一定的借鉴；三是通过对研究前沿的识别、遴选、解读，以数据和可视化图表的形式

把研究前沿领域展示给科研者、管理者和决策者，为其开展科学研究、制订科学技术发展的路线、战略和政策提供依据；四是完善图书馆的学科服务体系，为开展研究前沿分析服务、知识产权信息服务和学科发展分析服务等提供参考。

2 研究内容及方法（思路、方法、具体内容）

2.1 理论研究

自 1965 年 PRICE 将研究前沿概念引入科技领域，学者就其概念内涵、识别方法、数据来源等进行了多方面研究。

2.1.1 研究前沿概念内涵

研究前沿一般可分为基础研究前沿和应用研究前沿，前者又可称为科学研究前沿、研究前沿，以学术论文、研究报告等为数据源，后者又可称为技术研究前沿，主要指以专利及专利家族为主的数据源。目前研究前沿尚无明确统一的定义，不同的学者有不同的表述。详见表 1。由表 1 可知，在现有的研究中，学者通过不同的探测识别方法，得到了不同表现形式的研究前沿，从而给予研究前沿不同的定义。总体而言，关于研究前沿的定义可以分为三类：通过共被引聚类得到的高被引文献，通过文献耦合得到的施引文献和通过词语分析得到的突发或热点主题。

表 1 不同学者关于研究前沿表述一览表

年份	作者	主要观点
1965 年	PRICE, DJD	领域内的一组新颖且拥有高被引量的论文集合，高被引论文集合所反映的内容代表了领域的研究前沿
1973 年	Henry, Small	通过统计文献共被引频次来确定领域核心论文集，并在此基础上进行

		共被引聚类分析和网络分析来识别研究前沿
1991年	R. R. Braam 等	研究前沿是科学共同体所关注的一系列相关问题与概念。
1994年	E. Garfield	(活跃的)研究前沿一般包括互相关联的共被引文献对, 文献共被引的次数越多, 越有可能代表一个确定的或新兴的主题或子领域
1994年	O. Persson	研究前沿是某领域高被引文献簇的施引文献簇, 这些施引文献簇更能够代表当下的研究内容, 而高被引文献簇应是研究前沿的知识基础
2003年	S. Morris	研究前沿由引文耦合得到, 并通过时间线可视化研究指出前沿会随着科学问题的发展而产生或消亡
2006年	CHEN chaomei	研究前沿为突发词分析所得的研究主题
2008年	N Shibata 等	研究前沿可以采用新出现的并达到一定规模的文献集合来表征
2010年	S. P. Upham	研究前沿是被科学家高度关注的科学技术领域, 是政府、投资者认为具有潜力的高新技术
2014年	许晓阳等	最近出现、正在兴起的研究主题或研究领域
2016年	郑彦宁等	某段时间某个研究领域中最新出现、正在兴起并引起科学家高度关注的研究主题
2017年	冯佳	研究前沿是具有“高关注度”和“新颖性”这两个特点的研究主题

2.1.2 研究前沿识别方法

现有的研究前沿识别方法可分为专家判断法和计量分析法两大类, 前者以德尔菲法、专家咨询法、头脑风暴法等为主, 后者以引文分析法和词频统计法等为主。除了上述两大类, 随着数据库的不断完善、机器学习算法的兴起, 新的识别方法也不断出现, 详见表 2。由表 2 可知, 随着技术的发展, 相关研究的深入, 新的识别方法不断出现, 方法之间的相结合也成为使用较多的方法。

表 2 研究前沿识别方法一览表

类别	小类别	
专家判断法		通过不同形式的专家访谈、判读得到研究前沿(张雪等, 2022)
		德尔菲法(张雪等, 2022; 沙振江等, 2015)
		专家咨询法(S. Upham 等, 2010)
		头脑风暴法(张雪等, 2022; 沙振江等, 2015)
计量分析法		通过对科技文献的科学计量分析而识别出研究前沿
	引文分析法	通过对高被引文献进行分析而识别出研究前沿
		直接引用(R. Klavans 等, 2006)
		共被引分析(H. Small, 1973)
		文献耦合(S. Morris 等, 2003)
词频统计法	通过对低被引或零被引文献, 或新出现的文献, 对其相关关键词频次进行分析而识别出研究前沿	

		基于词频(突发词) (CHEN chaomei, 2006)
		基于共词 (J. Joung 等, 2017)
机器学习算法		随着大规模文本处理技术的发展, 机器学习的准确性和决策效率得以提高并应用于研究前沿识别领域 (张雪等, 2022)
		文本聚类算法 (Xu S 等, 2021)
		文本分类 (C. Lee 等, 2018; 李欣等, 2021)
		时间序列分析 (Xu S 等, 2019; 岳丽欣等, 2020)
综合方法		将上述至少两种方法相结合

2.1.3 研究前沿识别数据来源

数据的选取对研究前沿识别的效果有重要影响。在科学前沿识别的研究中, 科技论文、专利、基金项目、科技规划等数据源被广泛使用, 但在早期研究数据源上, 大多研究仅以论文或者专利数据为唯一数据源。袁建霞等 (2022) 指出科技前沿方向识别的数据源包括体现过去 (科技史料、研究论文、发明专利等)、现在 (研究计划、资助项目、科研进展等)、未来 (战略规划、前瞻报告、预测报告等) 不同时期科技发展的多种成果数据。近年来, 多源数据融合的前沿识别方法逐渐受到关注, 不同数据源之间的内在联系, 可体现出科学理论与现实技术相互融合、转化、协同发展的关系。

表 3 研究前沿表述一览表

	论文	专利	项目	规划
论文	王贤文等 (2015) 李小涛等 (2018) 凌张宾等 (2020)	N Shibata, 2011 许晓阳等, 2016 卢嘉悦等, 2021 任佳妮等, 2021 熊书玲等, 2020, 2022	李广建等, 2016	
专利		张婷等, 2018 高楠等, 2020		
项目	刘博文等 (2019)		王效岳等, 2017 张婧等, 2019	
规划	张庆芝等, 2018 白如江等, 2020		白如江等, 2017 周彦廷等, 2019	白如江等, 2017

2.2 实证研究

结合上述研究，以蚕桑产业为例，对所构建的研究前沿识别方法体系进行实证研究。

2.2.1 数据来源及分析方法

论文数据来源于 CNKI 数据库，专利数据来源于 Incopat 全球专利数据库，地理标志数据来源于国家知识产权局、全国地理标志农产品查询系统等。

2.2.2 论文分析

(1) 总体情况

在 CNKI 中，以“篇关摘=桑蚕 + 桑园 + 桑树 + 桑叶 + 桑果 + 桑白皮 + 桑枝 + 桑葚 + 桑椹 + 蚕桑 + 家蚕 + 蚕种 + 蚕蛹 + 蚕蛾 + 僵蚕 + 蚕沙 + 蚕茧 + 蚕丝 + 蚕茧 + 蚕房”进行检索，得到数据 100563 条。

(2) 研究热点分析

基于蚕桑研究主题分析可以发现，蚕桑品种研究是重中之重，其中涉及到新品种、家蚕品种、家蚕新品种、桑（树）品种等；其次为家蚕微粒子病，重点是综合防治、防治技术、防治效果等；蚕桑产业研究也是重点，涉及到产业发展现状、存在的问题、发展对策，产业发展思考等；家蚕养育也得到了学界重视，涉及到小蚕共育、人工饲料、种桑养蚕等；桑白皮具有较高的药用价值，其研究涉及到中医学、中药学、儿科学等；从区域来看，浙江、江苏、四川等省份是研究关注的重点区域。

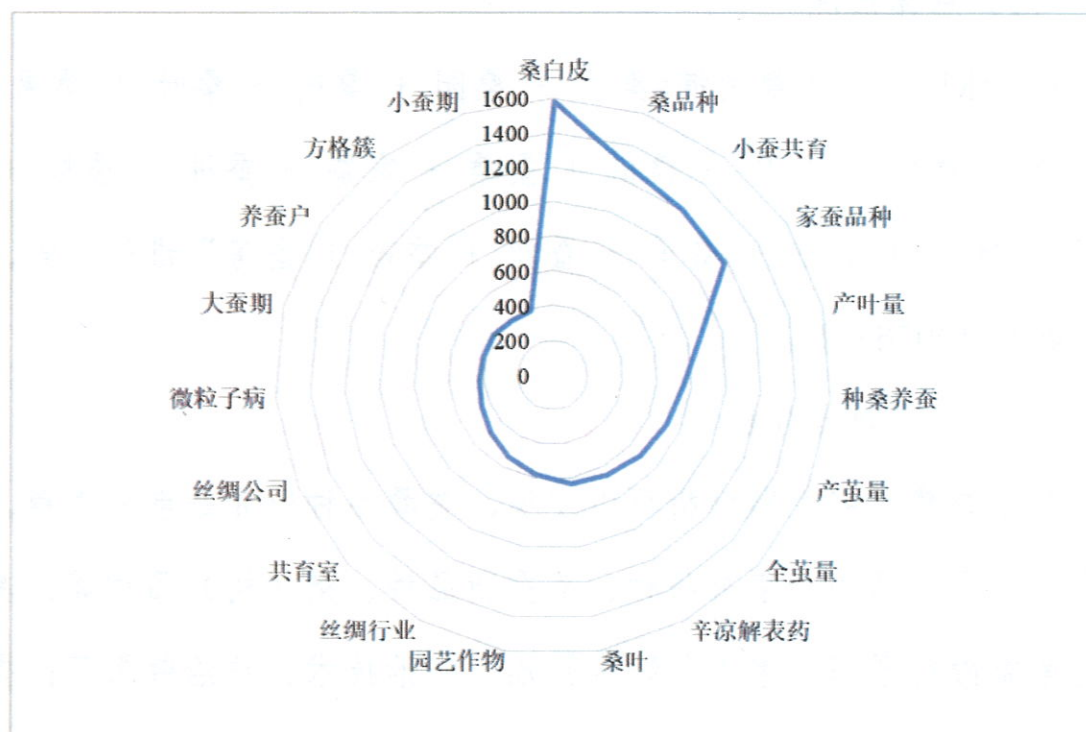
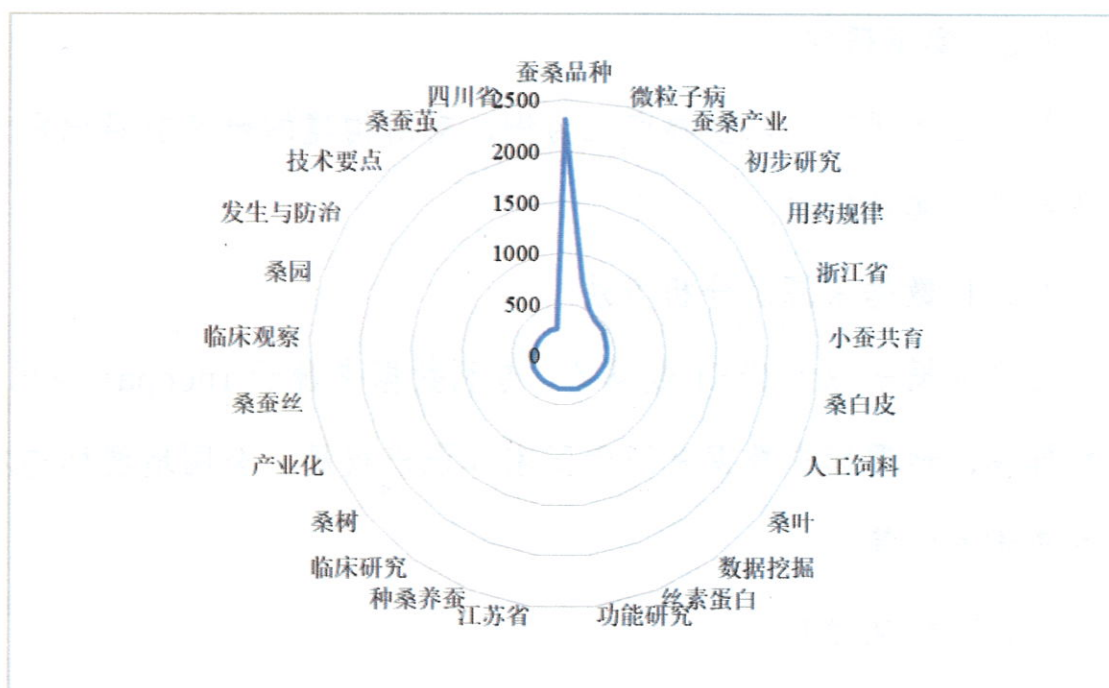


图 1 基于论文数据源的研究热点分析

(3) 研究机构分析

蚕桑研究机构主要为西南大学(含原西南农业大学)、苏州大学、中国农业科学院蚕业研究所、浙江大学、江苏科技大学、广西蚕业

技术推广总站、华南农业大学、云南省农业科学院蚕桑蜜蜂研究所、浙江理工大学、西北农林科技大学、山东农业大学、浙江省农业科学院蚕桑研究所、浙江农业大学、四川省农业科学院蚕业研究所、广东省农业科学院、安徽农业大学、广西大学、山东省蚕业科学研究所等。

2.2.3 专利分析

(1) 总体情况

在 Incopat 中，以“TIABC=桑蚕 or 桑园 or 桑树 or 桑叶 or 桑果 or 桑白皮 or 桑枝 or 桑葚 or 桑椹 or 蚕桑 or 家蚕 or 蚕种 or 蚕蛹 or 蚕蛾 or 僵蚕 or 蚕沙 or 蚕茧 or 蚕丝 or 蚕茧 or 蚕房”进行检索，得到数据共 96703 条，82338 个专利族。

(2) 技术领域

从 IPC 分类号来看，桑蚕专利主要集中在 A 部（人类生活必需）、C 部（化学；冶金）、D 部（纺织；造纸）和 B 部（作业；运输）。

桑蚕专利涉及的技术领域主要包括：A61K（医用、牙科用或梳妆用的配制品）；A61P（化合物或药物制剂的特定治疗活性）；A23L（食品、食料或非酒精饮料及其制备或处理，例如烹调、营养品质的改进、物理处理；食品或食料的一般保存）；A23K（专门适用于动物的饲养饲料及其生产方法）；A61Q（化妆品或类似梳妆用配制品的特定用途）；A23F（咖啡；茶；其代用品；它们的制造、配制或泡制）；A01K（畜牧业；养鸟业；养蜂业；养鱼业；捕鱼业；饲养或养殖其他类不包含的动物；动物的新品种）；A01G（园艺；果树栽培；林业）；

B32B（层状产品，即由扁平的或非扁平的薄层，例如泡沫状的、蜂窝状的薄层构成的产品）；C12G（药酒、保健酒等酒精饮料及其制备）等。

A部主要是桑蚕与人类生活必须相关的专利，涉及到医学、卫生学，食品科学，农业、林业、畜牧业等，技术领域：一是桑蚕用于医用配制品和传统草药，化妆品，蚕丝蛋白医用假体材料等，如桑叶、桑白皮、桑枝、桑椹/桑葚，蚕蛹、蚕蛾、僵蚕、蚕沙等用于医用配制品和传统草药，桑白皮、僵蚕用于化妆品；二是桑蚕用于食品或食料，如蚕蛹用于食品，桑椹用于饮料，桑叶用于保健茶，蚕蛹、蚕沙用于饲料等；三是家蚕饲养及桑蚕相关的作物栽培，如家蚕饲养方法与设备，蚕蛹用于虫草培植，桑枝、蚕沙用于食用菌栽培等；四是由桑蚕产生的家庭用的物品或设备，如床上用品及生产工艺，服装，保健枕等。C部主要是桑蚕与化学生物相关的专利，技术领域：一是与生物遗传相关的研究，如酶、突变或遗传功能、DNA，酶或微生物的测定与检测，蚕蛹用于基因遗传等；二是家蚕副产品的生物化学应用，如蚕蛹用于药酒，蚕蛾用于保健酒，蚕蛹用于虫草培植，以及蚕沙用于饲料、蚕沙堆肥装置及方法等。D部主要是桑蚕与纺织相关的专利，技术领域包括蚕丝生产，蚕丝蛋白面料，蛋白纤维制造等。B类主要是桑蚕制品领域的相关专利，涉及的技术领域为相关的薄层层状产品、一般过滤器等。

（3）申请区域、申请人分析

蚕桑领域的专利申请量排名前 10 位的分别是江苏、安徽、山东、浙江、广西、广东、四川、河南、北京、重庆等。

蚕桑领域的专利申请人类型主要为企业，其次依次为个人、大专院校、科研单位、机关团体、其它；主要申请人为西南大学、苏州大学、浙江大学、浙江理工大学、江苏科技大学、广东省农业科学院蚕业与农产品加工研究所、广西大学、东华大学、华南农业大学、四川省农业科学院蚕业研究所等。

2.2.4 地理标志分析

(1) 总体情况

蚕桑地理标志产品种类多样，达 88 种，其中地理标志证明商标 45 件，地理标志保护产品 19 个，农产品地理标志 24 个，涉及到“桑—蚕—茧—丝—绸”的全产业链条。

(2) 主要种类分析

蚕桑地理标志产品种类包括三个层面：一是与桑相关的地理标志，如桑树本身所生产的桑果、桑椹/桑葚、药桑等桑果实，桑叶及桑叶茶、桑叶鸡，桑树皮加工的纸品，桑基鱼塘衍生产品塘鱼和湖羊，以及桑黄；二是与蚕相关的地理标志，如蚕、桑蚕，蚕蛹及其衍生品；三是与茧、丝、绸、锦、绣、绢、纱有关的地理标志，如蚕茧、桑蚕茧，丝、蚕丝、茧丝、蚕丝被、丝绸，锦、绣、绢、纱等。

(3) 主要区域分析

从区域来看，地理标志产品分布于 19 个省份，其中四川拥有量最多，达到 20 个，其次为山东，有 12 个，江苏拥有 9 个，河北和辽宁分别拥有 6 个，上述 5 个省份占比达到 60%。

3 结论与建议

从基础研究、技术研发、产业发展三种不同角度对比论文、专利和地理标志等多源数据，可以发现蚕桑研究及产业发展呈现以下特点：一是从研究内容来看，论文主要关注蚕桑品种，家蚕微粒子病，蚕桑产业，家蚕养育，桑白皮等，专利主要关注蚕、桑、茧的应用，地理标志侧重于不同区域的蚕、桑、茧、丝、绸等特色发展，但都是贯通的，进行桑蚕良种繁育，可以带来优质的茧、丝、绸，并助力全产业链发展；二是从区域来看，江苏、山东、四川等是专利分布大户，也是地理标志分布的主要区域；三是从研究机构来看，西南大学（含原西南农业大学）、苏州大学等高校无论是基础研究还是技术研发都是主体，虽然企业是专利申请人主要类型，但具体到某个而言，表现不佳。

归纳起来，基于多数据源主题对比的科学研究前沿识别与基于论文或专利等单一数据源分析方法相比，能够克服单一数据源所存在的缺点，前瞻地识别出具有未来发展潜力的新兴科学研究前沿主题，并全面揭示相关产业发展的全貌。

4 项目成果（发表的文章、开发的软件、取得的实践效果等）

本研究形成 1 份主报告《蚕桑产业发展技术分析报告》和 3 份子报告《桑蚕基础研究分析报告》、《桑蚕专利技术发展分析报告》和《全国蚕桑类地理标志分析报告》，并形成 1 篇论文《基于论文和专利的桑蚕领域研发现状及前沿热点分析》（待投稿）。

5 参考文献

张雪, 张志强, 曹玲静, 阮伟南, 任晓亚, 冯志刚. 学科领域研究前沿识别方法研究进展[J]. 图书情报工作, 2022, (12): 139-151.

卢嘉悦, 李艳. 基于论文和专利数据的研究前沿挖掘研究——以智能网联汽车领域为例[J]. 中国发明与专利, 2021, (1): 13-20.

许晓阳, 郑彦宁, 刘志辉. 论文和专利相结合的研究前沿识别方法研究[J]. 图书情报工作, 2016, (24): 97-106.

秦婕, 李国强. 从专利技术现状看我国蚕桑业区域化发展[J]. 农业与技术, 2015, (11): 126-127, 136.

郭剑坤, 胡瑜. 我国蚕桑产业专利现状分析[J]. 中国蚕业, 2013, (4): 47-51, 59.

庞淑婷, 程光伟, 刘颖. 基于 incoPat 的中美高端纺织产业专利申请与转让分析[J]. 中国质量与标准导报, 2022, (6): 42-46.

晏宇杭, 周兰玉, 周永峰, 吴清华, 卢丽洁, 朱兴龙, 王越, 裴瑾. 基于 incoPat 专利数据库的白芷专利格局分析[J]. 中草药, 2021, (24): 7728-7738.