

基于文本挖掘的我国绿色金融政策研究

廖玉清

(上海理工大学 管理学院, 上海 200093)

摘要: 以我国 2017—2019 年绿色金融相关政策为数据源, 通过提取政策文件的文本—行特征词和高频关键词, 构建词云图、共现矩阵, 采用高频词社会网络分析等方法, 实现对这三年绿色金融相关政策内容信息的量化分析, 从政策制定侧重点及政策内容上对文本进行总结。结果表明, 我国 2017—2019 年绿色金融政策体现了政府对绿色金融企业机构激励机制及监管制度的完善, 体现出这三年绿色金融发展的侧重点。

关键词: 绿色金融; 文本内容挖掘; 词云; 共词分析

中图分类号: C 939 **文献标志码:** A **文章编号:** 1009-895X(2023)02-0219-08

DOI: 10.13256/j.cnki.jusst.sse.2023.02.017

Chinese Green Finance Policy Research Based on Text Mining

LIAO Yuqing

(Business School, University of Shanghai for Science and Technology, Shanghai 200093, China)

Abstract: With the relevant policies of green finance in China from 2017 to 2019 as the data source, the quantitative analysis of the content information of green finance related policies is conducted by extracting the text line feature words and high-frequency keywords in documents, constructing the word cloud map, co-occurrence matrix, high-frequency word social network, and summarizing the text from the policy formulation focus and policy content. The results show that during 2017-2019 China's policy of green finance improves the government's incentive mechanism and regulatory system for green finance enterprises and institutions, and reflects the focus of green finance development in the past three years.

Keywords: green finance; text mining; word cloud; co-occurrence analysis

自改革开放以来, 我国的经济得到迅速发展, 经济实力排名跻身世界前列。但是, 伴随我国经济的高速发展而来的是自然资源的大量开采以及生态环境的严重破坏。为应对气候变化并节约集约高效利用自然资源, 我国开始探索绿色经济的发展道路。金融作为现代经济的核心, 在社会资源配置中起着至关重要的作用。绿色金融能够引导资金从高耗能、

高污染行业逐步退出, 更多地流向绿色环保产业, 是推动经济可持续发展和产业升级转型的有力保障^[1]。

绿色金融的实质是围绕绿色投融资、项目和风控开展的一揽子金融产品、市场制度以及各种政府监管措施的统一有机结合体^[2]。目前我国环境质量任务繁重, 环保投资需求巨大, 构建绿色金融体系正当其时^[3]。通过对我国绿色金融相关政策及制度

收稿日期: 2020-08-18

作者简介: 廖玉清, 女, 博士。研究方向: 技术转移。E-mail: 13761747170@163.com

的梳理,可以看出我国已将绿色金融作为国家的战略目标,并通过出台各项政策,制定相关规定来保障国内绿色金融的发展^[4]。但与先进国家相比,法律法规体系建设滞后、财政资金补贴力度有限、政策性机构作用发挥不足等问题依然存在^[5]。因此,应激发地方政府推进绿色产业和绿色金融发展的积极性,并确保信息沟通的有效性和绿色金融政策的一致性,提升可测度的监管质量^[6],以完善绿色金融体系的建设。

近几年,伴随着“一带一路”政策以及乡村振兴战略的逐步推进,绿色金融的政策体系逐渐完善。本文主要对2017—2019年绿色金融相关政策进行分析研究,梳理在政策引导下的绿色金融产业的发展侧重点,以展示这三年绿色金融政策的发展格局。

一、研究回顾

基于政策解构视角,王凤荣等认为政府绿色金融政策主要分为两个维度的政府行为,一方面是来自经济领域的政府行为,包括财政项目拨款、税收优惠等经济手段,另一方面是来自于管控投资方向、绿色产品属性、生产数量限定等方面的行政手段和法律手段^[7]。Fischer认为绿色金融产品能够拉动社会资本投入绿色环保领域,但企业环境行为的信息披露不足使污染事件屡禁不止,不利于绿色金融发展^[8]。Ghisetti等认为良好的法律环境是绿色金融健康发展的重要保障,配套的法律制度不健全将会使绿色金融的效果大大降低^[9]。

在政策量化研究的领域中,文本挖掘已经得到广泛应用,尤其在海量的政策文本研究中产生了大量研究成果。现有研究主要是基于绿色金融投入主体的相关指标来测算绿色金融发展水平。Miao等认为绿色金融发展离不开中小企业市场的支持力量,但绿色金融理念却在中小企业中没有得到广泛应用,这在很大程度上抑制了经济发展转型和绿色金融产品创新^[10]。丁杰将双重差分法引入绿色信贷政策有效性研究,证明了绿色信贷抑制了重污染企业的信贷融资^[11]。在排污权交易方面,也有学者关注政策监管问题。黄韬等发现在现有的金融市场和金融法治格局下,监管权力分散、制度不完善等因素会导致绿色债券市场存在的监管套利问题^[12]。

随着计算机方法的引入和应用,政策文本分析所能处理的素材量和处理精度得到了大幅提升,并引入了新的方法和理念^[13]。文本挖掘技术可以从海

量政策数据中抽取隐含的知识,解读和获知政策的立场、倾向以及广义的政策比较分析^[14]。系统性、客观性、定理性是内容分析法的特点和优势,胡嫣然基于文本挖掘,采用内容分析法对我国铁路运输企业的财税支持政策进行量化分析和研究^[15]。赵公民等通过引入扎根理论和文本挖掘的研究方法,对政策文本进行词频分析、语义网络分析和中心性分析,研究广东省科技金融系统的运行模式^[16]。

通过梳理现有文献,发现大部分学者在绿色金融政策领域的研究为定性研究,对政策文本的定量研究缺乏。因此,本文以基于文本挖掘的文本可视化分析方法,对这三年我国绿色金融政策的发展侧重点进行梳理以呈现发展格局。

二、数据来源和处理方案

(一)数据来源

本文所分析的政策文本数据来源于“法律之星”网站收录的绿色金融政策文件,使用“八爪鱼”数据爬取工具,爬取2017—2019年绿色金融相关的政策文件,共获得这三年的政策文本数据量1509份,其中包括中央政策文件、部门规章以及各省市政策文件。笔者将得到的政策文本进行梳理,剔除重复文件,最终得到1487份绿色金融政策文本。

(二)文本处理方案

本文以2017—2019年我国绿色金融政策文件为实验文本数据,利用内容挖掘系统ROSTCM6和社会语义网络分析工具UCINET,通过文本预处理、文本高频词提取、高频词共现性分析、社会语义网络构建这四个步骤,得到可视化的高频关键词词云图和社会语义网络来分析这三年我国绿色金融政策的发展趋向和侧重点。其具体的实验步骤如下。

(1)文本预处理。由于收集到的文本数据包含噪声,须对文本进行预处理。首先,对文本进行去重,删除如“\”“?”“《”“》”等无意义符号,获得简单清洗后的文本。其次,由于收集到的文本数据为中文,故须进行中文分词。本文利用ROSTCM6提供的分词工具进行中文分词。最后,对分词后的文本进行去停用词操作,得到预处理后的文本数据。

(2)文本高频词提取。词汇频率不仅是文本的基本单位表征,而且反映了文本的语义和关注焦

点。文章中的高频关键词往往能反映出文章作者对某一领域的关注程度。本文利用 UCINET 工具对预处理后的文本数据进行词频统计分析，得到 2017—2019 年绿色金融政策文件的高频词汇。

(3) 高频词共现性分析。基于提取到的绿色金融政策文本数据中的高频关键词，统计同一时间维度下两个关键词之间的共现频率构建文本—高频词共现矩阵，结合统计到的高频关键词的共现矩阵来分析政策文件的核心内容。

(4) 社会语义网络构建。为了挖掘高频关键词间的联系，本研究利用 UCINET 工具结合 NETDAW 工具，经程序后台处理，自动绘制绿色金融政策文本中高频词共现关系矩阵的社会语义网络，以表现关键词间的内在逻辑。

三、绿色金融政策文本挖掘

(一) 分词提取及词频统计

首先，笔者在剔除重复文件之后，删除各发文机构、附件流程及人名等无实际意义的干扰词语；其次，由于现有的内容挖掘系统内置的自定义词表以及停用词表与本研究主题关联不大，为保证分析结果的有效性，本文以前期搜集到的政策文件解读为依据，比如将“绿色信贷”“信用信息”“碳金融”等词语添加至自定义词表，并利用 ROSTCM6 分词功能反复修改并更新自定义词表，以得到更准确的分词结果；最后，分别对这三年的政策文本进行分词、词频统计分析，得到共现关键词如表 1 所示。

表 1 绿色金融政策文本共现关键词部分例表
Tab. 1 Examples of co-occurrence keywords in green finance policy text

2017年		2018年		2019年	
关键词	词频	关键词	词频	关键词	词频
创新	6 327	创新	6 242	创新	5 172
农业	4 484	资源	4 984	资源	4 314
改革	4 082	科技	4 975	工业	3 928
金融机构	2 526	工业	4 127	监管	3 727
农产品	1 375	农业	3 722	环境	3 645
金融服务	1 263	生态	3 529	科技	3 502
养老服务	1 163	监管	3 526	安全	2 926
大数据	1 033	开发	3 409	农业	2 839
环境保护	998	金融机构	2 347	生态	2 737
小微企业	921	旅游	2 336	教育	2 621

(二) 共现性分析

为展现这三年政策文本关键词之间的联系，以词语之间的联系来表示政策文本的核心内容。统计两两关键词之间在同一年度的政策文本中共同出现的频次，构成文本—高频词共现矩阵，其频次越高，说明两两关键词之间的关系越紧密。但是，仅靠关键词之间的频次高低反映词语之间的联系紧密程度太过单薄，要进一步将原生共现矩阵转化为相似矩阵。相似矩阵中的数值范围在 [0,1]，数值越接近 1，

说明两个关键词的距离越近，关联程度越高。以 2017 年政策文本—高频词相似矩阵为例，如表 2 所示。

本研究进一步对整个关键词网络进行密度测量，密度的取值范围在 [0,1] 之间，密度越接近 1，表示整个关键词网络节点连接紧密且整个网络的复杂性较高。使用 UCINET 的密度测量工具对 2017—2019 年政策文本的网络密度进行测量，得出：这三年的网络密度值分别为 0.612 0，0.513 8 和 0.469 1，说明这三年的绿色金融政策文本具有较复杂的网络体

表 2 2017 年绿色金融政策文本部分高频词相似矩阵
Tab. 2 2017 green finance policy text high frequency words similarity matrix

高频词	创新	农业	环境	改革	保险	规划	保障	资金	监管	绿色	科技	经营
创新	1.000											
农业	0.439	1.000										
环境	0.943	0.466	1.000									
改革	0.943	0.466	1.000	1.000								
保险	0.500	0.887	0.530	0.530	1.000							
规划	0.884	0.481	0.938	0.938	0.547	1.000						
保障	0.986	0.433	0.929	0.929	0.493	0.901	1.000					
资金	0.928	0.459	0.985	0.985	0.522	0.955	0.944	1.000				
监管	0.841	0.522	0.892	0.892	0.594	0.920	0.829	0.879	1.000			
绿色	0.885	0.496	0.939	0.939	0.565	0.936	0.873	0.925	0.951	1.000		
科技	0.608	0.737	0.644	0.644	0.769	0.665	0.599	0.635	0.722	0.686	1.000	
经营	0.797	0.511	0.845	0.845	0.627	0.872	0.786	0.832	0.911	0.900	0.762	1.000

系，且关键词网络各节点之间的联系密切，文本信息内容主题较丰富；均值方差分别为 0.203 9，0.481 8 和 0.472 9，说明较 2017 年相比，2018 年和 2019 年的关键词网络中的主要关键词之间维系小范围的紧密程度，绿色金融政策发展的侧重点有明确的主题分布。

四、政策文本挖掘的社会网络分析

绿色金融政策文本的分析法主要有以下三种：（1）定性政策文本分析，多为对话语性和语义的分析，即从某一视角对政策文本资料中的某个词语进行解读和分析；（2）定量分析，通过定量来识别政策文本中可能出现的使用次数较多的关键词语，将其中的定义综合作为关键词或政策的核心关键词，进而不断分析和挖掘政策文本背后可能隐含的信息；（3）政策的综合分析，此方法主要集合了政策的定性分析与定量分析，除了对制定政策的文本内容进行主观的定性分析外，也同时包括定量的研究以及对未来政策的发展趋势的综合预判。

由于本研究的政策文本数据量较大，内容主题维度较高，为更好地呈现高频主题关键词的可视化效果，笔者先对政策文本进行提取行特征词操作，剔除意义不大的高频关键词，以从文中抽取的特征词进行量化来表示文本信息，如表 3 所示；再将 2017—2019 年的文本一行特征词分别导入到词云

图生成网站 WordArt，构建相应的词云图。

笔者将整理好的政策文本一行特征词导入到词云图生成网站 WordArt，其中在特征词文本中出现频次最高的关键词在词云图中越突显。使用“UCINET→NETDAW”功能分别构建这三年的绿色金融政策文本的社会语义网络，其中节点表示关键词，连线表示关键词之间的共现关系。

表 3 2017—2019 年绿色金融政策部分文本一行特征词
Tab. 3 2017-2019 green finance policy part text-line words

2017	创业 创新 住宅 强化 直接融资 节能 绿色 污染 信息共享
	政府引导 创新 强化 节水 知识产权 绿色 污染 信息共享 消费
	融资服务 定位 改革创新 保障 集聚 保险 保护 金融业 水环境
2018	财政 完善 创新 投资 节能 绿色 污染 中小企业 服务
	认证 鼓励 贷款 协调 融资 环保 环境 监测 监管 监测 监管 审批 绿色发展 绿色金融 培训 基金 中国 改革
2019	住房 农业农村 银行 环境 监管 绿色发展 绿色金融 改革 生态
	住房 绿色 信息共享 服务 农业农村 银行 新能源 节能环保 融资
	环境 监管 绿色建筑 动态 基金 改革 生态 数字 资源

（一）2017 年绿色金融政策的词云图与社会网络分析

2017 年我国绿色金融政策已具有较清晰的基础架构。从图 1 所示的 2017 年绿色金融政策文本词

云图可以看出,“创新”“互联网”“大数据”“科技”“监管”“评估考核”“金融机构”“保险”“信贷”“基金”“担保”“教育”“农业”等词语在词云图中均有突显。结合图2所示的2017年绿色金融政策社会语义网络可知:随着信息技术的发展成熟,绿色金融体系也要适应社会的发展,重视改革创新;在之前的绿色信贷、绿色保险、绿色债券以及绿色中介服务等主体绿色金融产品的基础上,更多地强调利用互联网、大数据技术创新绿色金融产品的重要性;在绿色金融发展中政府发挥着重要作用,政策性金融比传统金融更需要构建科学的约束机制,提高参与绿色金融的产业机构的监管



图1 2017年绿色金融政策文本词云图

以及信用信息共享能力,加强对绿色金融产业参与资历和业绩的评估考核力度。此外,2017年“乡村振兴”战略的提出,使得我国绿色金融政策加大对农村农业绿色发展的融资扶持,发展绿色金融助推乡村振兴,国土资源配置等得到改善。

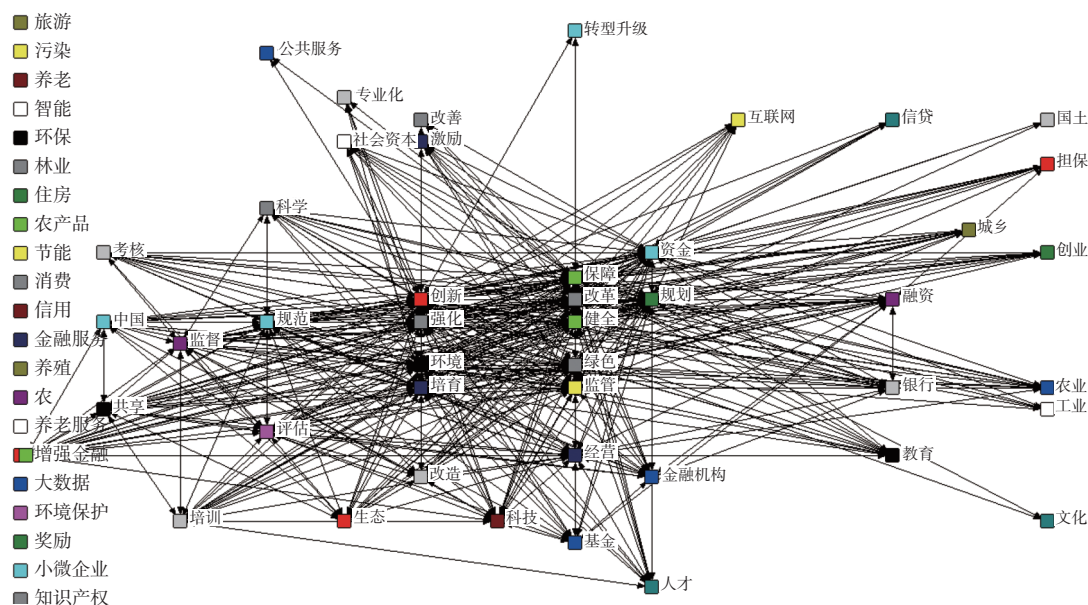


图2 2017年绿色金融政策社会语义网络

Fig. 2 2017 green finance policy social semantic network

(二) 2018年绿色金融政策的词云图与社会网络分析

从图3所示的2018年绿色金融政策词云图可以看出,“创新”“服务”“优化”“资源”“协调”“环境”“审批”“监管”“考核”“投资”“扶持”“培育”“保险”“农业”等词语在词云图中均被突显出来。图4的2018年绿色金融政策社会语义网络说明:2018年的政策在2017年的基础上没有增加关于绿色金融产品的新内容,但是政府的引导性地位更加突出,笔者认为各级绿色金融体系应保持创新的前进动力,加大对绿色金融企业机构的投资力度,扶持中小微企业的绿色转型,要求参与绿色金融的企业机构提供优质的金融服务;

注重环境与生产的协调性,增强绿色保险的突出作用,强调环境污染责任保险工作的重要性;完善企业或者机构参与绿色金融的审批机制,优化绩效考核体系;注重发展农村绿色农业,促进绿色生态示范区的建设;鼓励创新培训,金融人才是绿色金融体系不断专业化、多元化的生力军。



图3 2018年绿色金融政策词云图

Fig. 3 2018 green finance policy word cloud chart

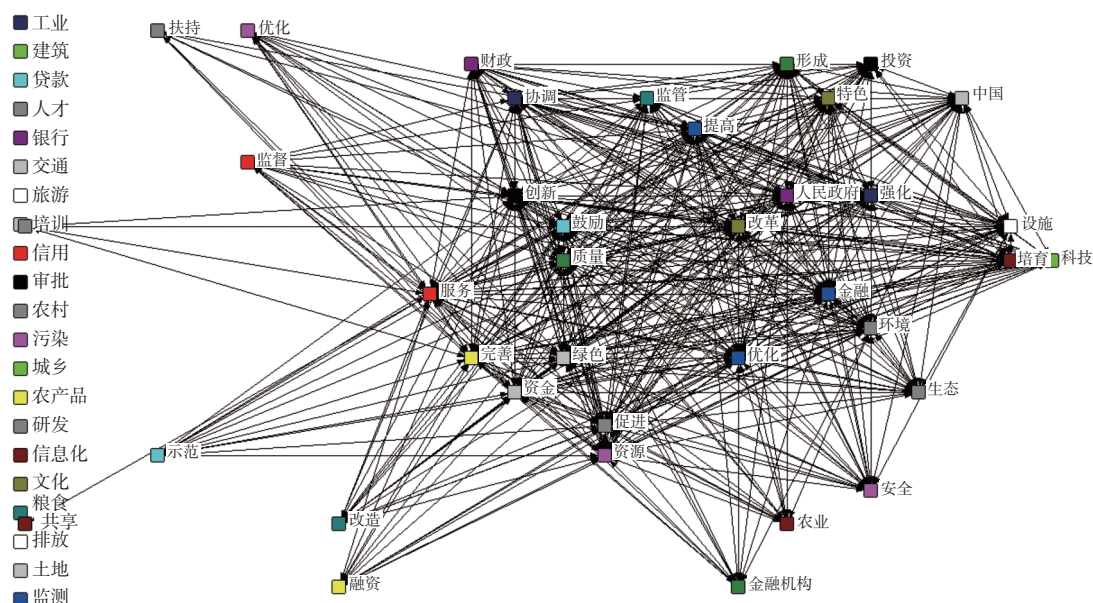


图4 2018年绿色金融政策社会语义网络

Fig. 4 2018 green finance policy social semantic network

(三) 2019年绿色金融政策的词云图与社会网络分析

从图5所示的2019年绿色金融政策词云图可以看出,“创新”“资源”“服务”“监管”“环境”“科技”“培训”“金融机构”“教育”“公共服务”“高质量”“农业农村”“农产品”“知识产权”“银行”“共享”等词语在词云图上均被突显出来。结合图6的2019年绿色金融政策社会语义网络可知:和2017年、2018年相比较,2019年的政策体现出绿色金融作为一种政策性金融比传



图5 2019年绿色金融政策词云图

Fig. 5 2019 green finance policy word cloud chart

统金融更注重社会效益的特征,且政策针对性更强;坚持利用好新科技工具如大数据、互联网、人工智能不断创新升级绿色金融体系,建立健全并参与绿

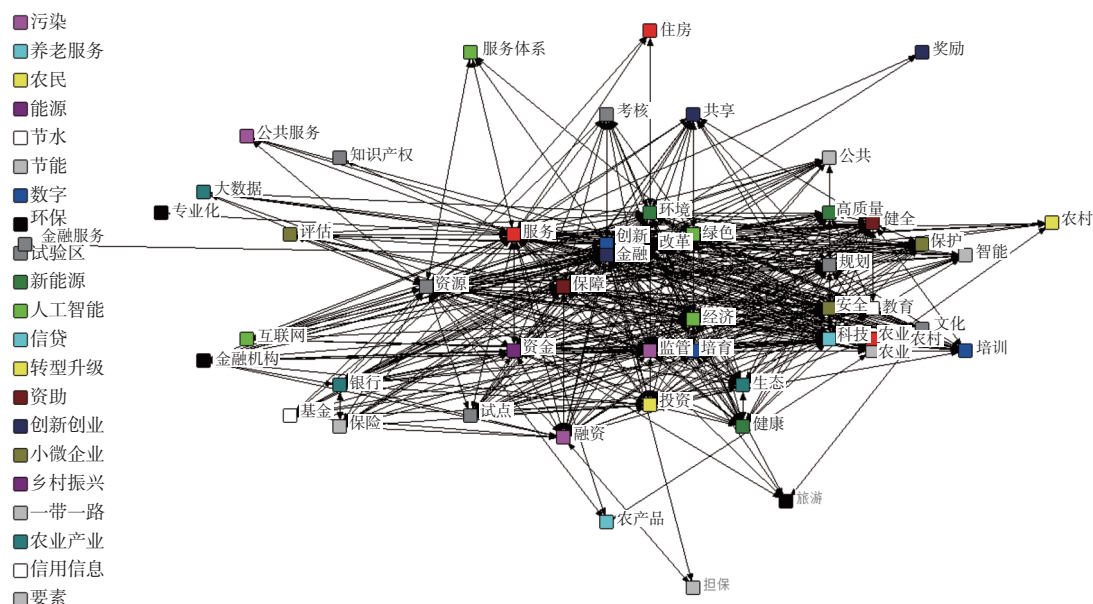


图6 2019年绿色金融政策社会语义网络

Fig. 6 2019 green finance policy social semantic network

色金融企业机构的信息共享平台,增强互联效应;加强政府监管力度,并促进社会督促约束绿色金融企业机构加快完善金融服务体系;随着各种IP热的兴起,绿色金融开始重视对知识产权的绿色转型;保障农村农业的绿色金融的融资力度,发展农产品以及文化旅游的品牌营销,切实为乡村振兴战略做出贡献。

五、结语

本文对2017—2019年我国绿色金融政策进行梳理分析,探寻当前绿色金融政策体系的发展侧重点,以词云图结合社会语义网络展现我国这三年的绿色金融政策体系发展格局,主要结论如下。

一方面,由上述2017—2019年词云图与社会语义网络的对比分析可知,在政策制定上,每一阶段的绿色金融政策制定都是随着不同的时代发展背景而变化。2017年和2019年的政策文本高频关键词以及社会语义网络以“创新”“监管”“服务”

“绿色基金”“农村农业”等词语为核心词,既有政策引导的关键词,也有对绿色金融发展体系中特定绿色金融产品的针对性措施关键词,而2018年的政策文本的高频词以及社会语义网络主要以“协调”“创新”“监管”等政策引导词为核心词,没有对绿色金融产品有明确的发展策略关键词,更多地是为2017年巩固政策,为2019年作政策铺垫。

另一方面,在政策内容上可知:第一,绿色金融作为一种政策性金融,政府在其体系的构建以及发展中起着主要作用,这三年的政策文件充分显示了政府对参与绿色金融的企业机构的激励机制以及监管制度的制定和完善,引导金融机构对绿色金融加大投入力度,提供更完善的绿色金融服务,建立更完善的绿色金融项目清单;第二,我国绿色金融政策体系的基本架构已经形成,依据全国各地的发展水平以及信息技术发展水平将绿色金融建设成具有中国特色的发展体系,打造绿色企业的信用信息共享平台,加强各地互联,找寻社会新兴发展动力,比如知识产权、人工智能等,为我国的绿色金融体系注入新的活力;第三,由于“乡村振兴”战略的提出,农村农业的转型升级开始加入绿色金融体系的建设,国家发展绿色农产品品牌战略,发展文化旅游、体育旅游等特色绿色旅游产品;第四,加大了对高校以及科研院校的扶持力度,重点培养

高端金融人才,为绿色金融体系的长久发展提供专业活力。

本文对我国2017—2019年绿色金融政策的发展格局进行梳理,但由于数据的限制无法分析每年全部的政策文本,并且没有对政策进行更细致的分类分析,缺少政策发文机构之间的对比分析,在接下来的研究中将继续推进对绿色金融政策文本多角度的研究。

参考文献:

- [1] 邹锦吉. 绿色金融政策、政策协同与工业污染强度——基于政策文本分析的视角[J]. 金融理论与实践, 2017(12): 71-74.
- [2] 秦雨桐, 王静. 中国绿色金融发展现状及问题研究[J]. 天津经济, 2019(12): 20-25.
- [3] 环境保护部. 构建绿色金融政策体系 推动环境质量改善[EB/OL]. (2016-09-05) [2020-02-16] https://www.mee.gov.cn/gkml/sthjbgw/qt/201609/t20160903_363532_wh.htm.
- [4] 杨冠雄, 李吉祥, 刘红. 我国绿色金融法规政策构建的思考[J]. 西部金融, 2017(12): 90-92.
- [5] 周中明. 我国绿色金融政策支持体系构建[J]. 合作经济与科技, 2017(17): 54-56.
- [6] 贵斌威. 绿色金融发展的法律政策研究[J]. 知识经济, 2019(21): 65-72.
- [7] 王凤荣, 夏红玉, 李雪. 中国文化产业政策变迁及其有效性实证研究——基于转型经济中的政府竞争视角[J]. 山东大学学报(哲学社会科学版), 2016(3): 13-26.
- [8] FISCHER C. Environmental protection for sale: strategic green industrial policy and climate finance[J]. *Environmental and Resource Economics*, 2017, 66(3): 553-575.
- [9] GHISETTI C, QUATRARO F. Green technologies and environmental productivity: a cross-sectoral analysis of direct and indirect effects in Italian regions[J]. *Ecological Economics*, 2017, 132: 1-13.
- [10] MIAO C L, FANG D B, SUN L Y, et al. Natural resources utilization efficiency under the influence of green technological innovation[J]. *Resources, Conservation and Recycling*, 2017, 126: 153-161.
- [11] 丁杰. 绿色信贷政策、信贷资源配置与企业策略性反应[J]. 经济评论, 2019(4): 62-75.

- [12] 黄韬, 乐清月. 中国绿色债券市场规则体系的生成特点及其问题[J]. 证券市场导报, 2018(11): 41-49;58.
- [13] 裴雷, 孙建军, 周兆韬. 政策文本计算: 一种新的政策文本解读方式[J]. 图书与情报, 2016(6): 47-55.
- [14] 张骁, 周霞, 王亚丹. 中国科技服务业政策的量化与演变——基于扎根理论和文本挖掘分析[J]. 中国科技论坛, 2018(6): 6-13.
- [15] 胡嫣然. 基于文本挖掘的中国铁路运输企业财税支持政策研究[D]. 北京: 北京交通大学, 2016.
- [16] 赵公民, 刘金金, 武勇杰, 等. 基于扎根理论和文本挖掘的广东省科技金融政策共词网络研究[J]. 科技管理研究, 2019, 39(3): 51-57.
- (编辑: 程爱婕)

(上接第218页)

- [12] 孙荣华, 张建民. 基于创业生态系统的众创空间研究: 一个研究框架[J]. 科技管理研究, 2018, 38(1): 244-249.
- [13] 冒乔玲, 许敏. 技术创新驱动企业成长的绩效分析——基于创新型上市公司的实证研究[J]. 企业经济, 2012, 31(4): 17-23.
- [14] 陆小成. 低碳科技公共服务平台的功能与机制研究[J]. 城市观察, 2011(4): 126-133.
- [15] 马凤岭, 陈颀. 基于扎根理论的孵化器商业模式演进机制研究[J]. 科学学与科学技术管理, 2014(5): 130-136.
- [16] 雷善玉, 王焕冉, 张淑慧. 环保企业绿色技术创新的动力机制——基于扎根理论的探索研究[J]. 管理案例研究与评论, 2014, 7(4): 283-296.
- [17] 项国鹏, 俞金含, 黄玮. 科技企业加速器运营机制国际经验及对我国的启示[J]. 科技进步与对策, 2016, 33(20): 30-36.
- [18] 吴卫, 银路. 巴斯德象限取向模型与新型研发机构功能定位[J]. 技术经济, 2016, 35(8): 38-44.
- [19] 王雪原, 王宏起, 张立岩. 不同类别区域创新平台的功能定位及其协同发展研究[J]. 科技进步与对策, 2013, 30(15): 36-40.
- (编辑: 程爱婕)