

# 国家高新技术产业开发区生物医药产业政策主题挖掘及量化评价研究<sup>Δ</sup>

乔 晗\*,徐俐颖,李 旭,李树祥,褚淑贞<sup>#</sup>(中国药科大学国家药物政策与医药产业经济中心,南京 211198)

中图分类号 R95;F426 文献标志码 A 文章编号 1001-0408(2020)22-2689-06

DOI 10.6039/j.issn.1001-0408.2020.22.01

**摘 要** 目的:为我国生物医药产业发展提供参考。方法:收集2010年1月—2019年12月我国全部169家国家高新技术产业开发区(以下简称“国家高新区”)官方网站上发布的政策文件,从中筛选出专门针对生物医药产业方面的政策以及全部产业政策中提及生物医药产业发展的政策,采用Excel 2019软件进行政策信息的汇总分析;使用Gensim包完成政策文本的预处理并构建隐含狄利克雷分布(LDA)主题模型,实现对政策文本中潜在语义信息的提取与分析。结果:共收集到相关政策文本518件,其中有效词汇共58 617个,高频词汇包括项目、人才、支持、补贴、创新等;提取出8个主题,按主题强度排序依次为技术创新、人才建设、融资支持、项目金融支持、税收优惠、资源引领、出口贸易、中小企业建设,强度分别为0.299、0.168、0.134、0.116、0.113、0.063、0.058、0.049。结论:国家高新区生物医药产业政策以技术创新、人才建设和融资支持为主,缺乏对中小企业建设、资源引领、出口贸易的关注。今后国家高新区对生物医药产业的政策支持及体系建设应重点完善政策激励措施,平衡各领域政策运用比例;鼓励企业扩大海外市场,学习海外先进技术,加强科研等领域的多边合作;配套相应的政策以增进企业间的交流合作,发挥龙头企业和名牌产品的带动作用,促进中小企业发展。

**关键词** 隐含狄利克雷分布主题模型;国家高新技术产业开发区;生物医药产业;政策;主题挖掘;量化评价

## Study on the Subject Mining and Quantitative Evaluation of Biomedical Industry Policy in National High-tech Industrial Development Zone

QIAO Han, XU Liying, LI Xu, LI Shuxiang, CHU Shuzhen (Research Center of National Drug Policy & Ecosystem, China Pharmaceutical University, Nanjing 211198, China)

**ABSTRACT** OBJECTIVE: To provide reference for the development of biomedical industry in China. METHODS: The policies and documents issued by 169 national high-tech industrial development zone (hereinafter referred to as “National high-tech zone”) official websites from Jan. 2010 to Dec. 2019 were collected to screen policies specifically for the biomedical industry and the development policy of biomedical industry mentioned in all industrial policies. Excel 2019 software was used to summarize and analyze policy information. Gensim package was used to preprocess the policy texts and construct LDA topic model to extract the potential semantic information from the policy texts. RESULTS: A total of 518 policy texts were collected, including 58 617 effective words; high-frequency words included project, talent, support, subsidy, innovation, etc.; 8 themes were extracted, ranked by theme intensity as technological innovation, talent construction, financing support, project financial support, tax incentives, resource guidance, export trade, and construction of small and medium-sized enterprises, with the intensities of 0.299, 0.168, 0.134, 0.116, 0.113, 0.063, 0.058 and 0.049, respectively. CONCLUSIONS: The biomedical industry policy of the national high-tech zone is mainly based on technological innovation, talent construction and financing support, and lacks attention to the construction of small and medium-sized enterprises, resource guidance, and export trade. In the future, the national high-tech zone’s policy support and system construction for the biomedical industry should focus on improving policy incentives and balancing the proportion of policy applications in various fields; encourage enterprises to expand overseas markets, learn overseas advanced technologies, and strengthen multilateral cooperation in scientific research and other fields; support the corresponding policies to enhance exchanges and cooperation between enterprises, give play to the leading role of leading enterprises and famous brand products, and promote the development of small and medium-sized enterprises.

**KEYWORDS** LDA topic model; National high-tech industrial development zone; Biomedical industry; Policy; Theme mining; Quantitative evaluation

Δ 基金项目:国家社会科学基金重大项目(No.15ZDB167)

\* 硕士研究生。研究方向:医药经济。E-mail:qh\_4869su@163.com

<sup>#</sup> 通信作者:教授,硕士生导师。研究方向:医药产业经济与政策。E-mail:csz77844@163.com

国家高新技术产业开发区(以下简称“国家高新区”)作为改革开放以来我国经济体制改革的重要成果之一,是我国高新技术产业孵化的重要基地。1988年经国务院批准,我国建立了首个高新区——中关村高新

区;经过30多年的发展,国家高新区的队伍和规模不断发展壮大,截至2019年12月,我国国家高新区总数已达169家,完成了“示范、引领、辐射、带动”创新发展的政策目标,已成为推动国家创新发展的主力军<sup>[1]</sup>。近年来,生物医药产业作为国家拟定的战略新兴产业步入了高速发展阶段,逐渐成为了多个国家高新区发展的重点产业<sup>[2]</sup>。其中,以泰州医药高新区、上海张江药谷为代表的主导生物医药产业的园区共110家,占高新区总数的65.1%<sup>[2]</sup>;国家医药产业工业总产值从2007年的2 231.99亿元(占比6.66%)增长至2018年的8 495.38亿元(占比15.14%)<sup>[3-4]</sup>,在国家高新区的发展过程中占有举足轻重的地位。

科技园区是推动高新技术产业高速发展的重要载体,世界各国纷纷建立了不同发展模式的科技园区(如著名的美国硅谷科技园)以提升自身的国际竞争力<sup>[5]</sup>。其中,美国等西方国家的科技园区建设以市场为导向<sup>[6]</sup>,而我国高新区的建设与此不同,政府主导的科技园区建设决定了国家高新区的高速发展离不开政府强有力的产业规划能力,政策环境是高技术产业进行技术创新、人才引进和科学发展的前提条件<sup>[7]</sup>。政策作为推动我国国家高新区发展的重要工具,引起了众多学者的高度关注,目前学术界对于高新区政策文本的研究多以量化评价为主。由于政策文本是一种较为特殊的文本,包含较多的政策用词,政策用词的高维性在政策文本挖掘分析过程中会降低文本聚类的效果,因此有必要对政策文本进行语义分析<sup>[8]</sup>。隐含狄利克雷分布(Latent Dirichlet allocation, LDA)主题模型是一种非监督学习模型,能够提取文本中潜在的主题<sup>[9]</sup>。与传统的定性分析方法不同, LDA模型可以在政策文本聚类及主题提取的过程中避免人工编码的主观因素<sup>[9]</sup>。已有多位学者将LDA主题模型运用于政策文本的分析并取得了理想的实践结果,如郎政<sup>[10]</sup>利用网络爬虫软件收集了2006—2016年甘肃省14市(州)政府网站中体现政府职能的政策文本,并基于主题模型LDA算法分析中央和地方职能的匹配性和对应性,初步证实了该省经济发展动力不足、区域创新能力不足都与政府职能的匹配性有着很大的关联。本研究以与生物医药产业相关的国家高新区园区层级政策文本为研究对象,通过探索政策文本高频关键词和主题分布特点,从整体上把握国家高新区生物医药产业的政策主题及发展重点,以期为我国生物医药产业发展提供参考。

## 1 LDA主题模型的构建

LDA主题模型是一个基于三层贝叶斯拓扑结构的概率生成模型,其三层拓扑结构图见图1。该模型基于如下假设:(1)文档集合中存在 $K$ 个主题且主题之间相互独立;(2)每个文档由 $K$ 个主题随机混合组成,且主题

参数服从Dirichlet分布;(3)每个主题是特征词上的多项分布,该多项分布的参数服从Dirichlet分布。LDA主题模型广泛用于提取文本中潜在的语义信息,并将文本数据分为文档层、主题层和特征词层,旨在从概率的角度构建文档的生成过程<sup>[11]</sup>,其概率模型图见图2。

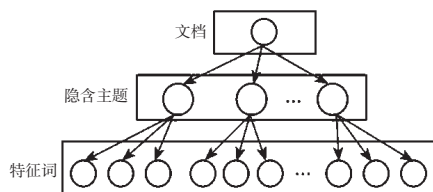


图1 LDA主题模型三层拓扑结构图

Fig 1 Three-layer topology structure diagram of LDA topic model

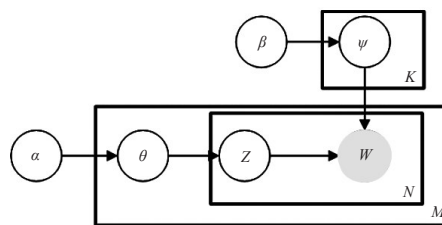


图2 LDA概率模型图

Fig 2 LDA probability graph model

图2中, $\alpha$ 、 $\beta$ 为先验参数,可根据经验给定; $K$ 为主题总数; $\theta$ 、 $\psi$ 分别表示每个文档的主题分布及每个主题的词分布(即每个主题中所包含的主题词及其相关度),是未知的隐含变量; $Z=(z_1, z_2, \dots, z_k)$ 表示文档的主题, $W=(w_1, w_2, \dots, w_r)$ 表示文档中的特征词,是可观测到的已知变量, $N$ 为特征词的个数。LDA模型生成文档的过程分为两步:(1)从Dirichlet( $\alpha$ )分布中生成每一篇文档的主题分布 $\theta$ 并生成主题 $Z$ 所包含的特征词 $W$ 。(2)从Dirichlet( $\beta$ )分布中生成 $K$ 个主题的词分布并据此生成特征词 $W$ 。以上过程重复 $M$ 次即可生成整个语料库。所有变量的联合分布计算公式如下:

$$P(W|D) = \sum P(W|Z) \cdot P(Z|D) \dots\dots\dots (1)$$

式中, $P(W|Z)$ 是主题 $Z$ 上的特征词分布值, $P(Z|D)$ 为文档 $D$ 上的主题分布值,两者的联合概率分布 $P(W|D)$ 构成了文档-主题-特征词关系。在LDA主题模型中, $\theta$ 和 $\psi$ 的参数估计可以使用变异期望最大化(EM)算法<sup>[12]</sup>和Gibbs采样<sup>[13]</sup>等统计技术来完成。

近年来,随着计算机技术的发展,已有多个软件可用于本文数据分析中的主题建模和挖掘,如SAS Text Miner和SPSS Clementine等商业软件包及R语言、Python和Java等程序包含的文本挖掘开源工具包<sup>[14]</sup>。Gensim包作为Python的第三方开源包,包含多种常见的自然语言处理模型,如词向量模型、TF-IDF(Term frequency-inverse document frequency)算法、LDA主题模型等。

鉴于Gensim包提供的自然语言处理工具在文本挖掘中应用的方便性及全面性,本研究采用Gensim包对国家高新区发布的生物医药产业相关政策文本展开研究。

## 2 资料来源

收集2010年1月—2019年12月我国现有169家国家高新区官方网站上发布的政策文件,从中筛选出专门针对生物医药产业方面的政策以及全部产业政策中提及生物医药产业发展的政策。排除仅体现政府对国家高新区发展态度的政策文本、与国家高新区相关性较低的政策文本(如一些人事变更通知等)等。文件类型包括法律、条例、意见、通知、办法、措施、细则、决定、方案等。采用Excel 2019软件进行政策信息的汇总分析。

## 3 文件检索结果及数据处理

### 3.1 文件检索结果及分布情况

按照上述政策检索和筛选方法,本研究共获得国家高新区生物医药产业相关政策文本518件,涉及高新区132家,其余37家国家高新区因其官网未发布涉及生物医药产业的政策文件或未建立官方网站而未获得有效信息。518件政策文本的地区分布见图3。由图3可知,与生物医药产业相关的政策文本主要集中在东部地区。

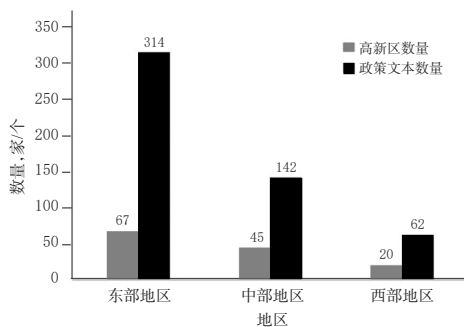


图3 518件政策文本的地区分布

Fig 3 Regional distribution of 518 policy texts

### 3.2 政策文本预处理

以518件国家高新区生物医药产业园区层级政策文本为语料库,采用Gensim中的Jieba分词工具包并结合正则表达式(去除字母、数字等非中文字符)进行中文分词。融合《中文停用词库》《哈工大停用词》作为本研究的停用词表,去掉长度小于2的词汇以及“企业”“组织”等政策文本的常规名词词汇,同时加载基于国家高新区和政策文本的自定义词典以提升分词效果,最终获得有效词汇58 617个。

为了构建合适的数据结构,使其具有Gensim主题建模可以处理的输入格式,本研究采用词袋模型(Bag of words)的形式来表示文档。该模型可以忽略每个词汇出现的顺序,将每篇文档表示成一个长向量。同时,引入TF-IDF算法进行词向量加权,建立词项文档矩阵(DTM)。TF-IDF算法可减少在多篇文档中频繁出现的

词汇的权重,并可提升稀有词汇的权重<sup>[15]</sup>,其计算公式如下:

$$TF_w = \frac{\text{在某一类中词条 } w \text{ 出现的次数}}{\text{该类中所有词条数目}} \quad (2)$$

$$IDF = \lg\left[\frac{\text{语料库的文档总数}}{\text{包含词条 } w \text{ 的文楼数} + 1}\right] \quad (3)$$

$$TF-IDF = TF \cdot IDF \quad (4)$$

### 3.3 模型参数设置

LDA主题模型取决于3个参数:Dirichlet超参数 $\alpha$ 和 $\beta$ 以及主题数量 $K$ 。本研究参考Griffiths TL等<sup>[13]</sup>提出的方案,将经验值 $\alpha$ 设为 $50/K$ 、 $\beta$ 设为0.1,作为初始模型参数;在初始参数值附近不断调试,同时观察模型结果确定最优参数,最终调参结果为 $\alpha=0.2$ 、 $\beta=0.1$ 。本研究采用“主题一致性(Topic coherence)”<sup>[16]</sup>评价主题模型,一致性分数(Coherence value)越高,表示主题可解释性越强。因本研究政策文本样本量为518,当主题数量过高时,模型不稳定,须结合主题一致性折线图(如图4所示)取10以下的最优主题数<sup>[17]</sup>。由图4可知,最优主题数 $K=8$ 。

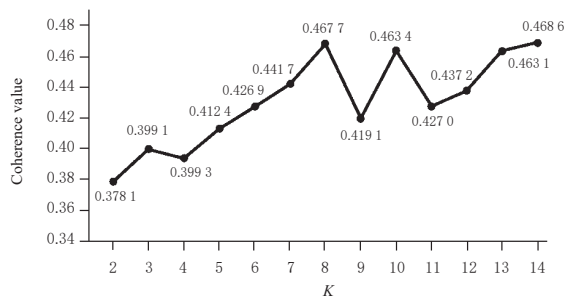


图4 主题一致性分数折线图

Fig 4 Thematic consistence value line chart

## 4 结果

### 4.1 高频词汇分析

采用上述方法对本研究纳入的政策文本进行预处理,并对分词得出的58 617个词汇进行词频统计,结果见表1(因篇幅所限,本文仅列出前10个高频词汇);选取排名前200位的关键词绘制词云图,结果见图5(图中关键词字号越大,词频越高)。

表1 政策文本高频词汇(前10个)

Tab 1 High-frequency words of policy texts (Top 10)

高频词汇	词频,次	占比,%	高频词汇	词频,次	占比,%
项目	2 441	4.16	资金	1 443	2.46
人才	2 319	3.96	补助	1 295	2.21
支持	2 071	3.53	创业	1 232	2.10
补贴	1 542	2.63	科技	1 114	1.90
创新	1 526	2.60	机构	1 104	1.88

上述政策文本关键词词云图基本可以体现我国国家高新区生物医药产业政策的集中情况。例如,高频词“支持”“扶持”等词汇体现出我国国家高新区生物医药



图5 政策文本关键词词云图

Fig 5 Policy text keyword cloud map

产业政策主要是以支持和激励为导向,引导企业发展。生物医药产业的特点是长周期、高投入、高风险、高利润,因此企业前期的资金风险较高,特别是中小型生物医药企业面临的处境更为困难,企业发展离不开资金,图中“补贴”“资金”等词汇体现出我国政策的侧重点在对企业的财政支持上,例如,北京中关村管委会发布的《中关村国家自主创新示范区天使投资和创业投资支持资金管理办法》<sup>[19]</sup>对企业成果转化类项目及产业转化类项目给予资金支持,带动了园区内企业协同创新平台的建设。如今,人才已成为各行各业发展不可或缺的资源,从图中也可看出,“人才”“团队”“补助”等词汇体现出园区政策对国家高新区人才建设的大力支持,如泰州医药高新区管委会发布的《关于加强中国医药城专业化高层次人才队伍建设的实施意见》<sup>[19]</sup>对医药城高层次创业人才给予资金奖励及生活津贴,同时解决其配偶就业问题,并提供子女入学保障、医疗保险、住房补贴、社会保障等完善措施,以解决创新创业人才的家庭顾虑。创新是生物医药产业发展的核心竞争力,激励生物医药产业创新发展已成为产业转型升级的重要方式,图中“创新”“科技”“知识产权”即体现出政策的关注点集中于企业创新,如泰州医药高新区管委会印发《关于推进科技创新引领高质量发展的若干政策》<sup>[20]</sup>,通过鼓励企业自主创新、推动科技成果转移转化、加强研发机构和载体建设多个渠道推动生物医药产业科技创新的高质量发展。

## 4.2 政策文本主题分析

4.2.1 主题强度分布 LDA 主题模型可以计算整个语料库中每个主题的比例,即主题强度,以反映每个主题

在语料库中的相对分量。由 LDA 主题模型训练得出每篇文档对应主题的后验概率以及每个主题对应主题词的后验概率,进而得出主题强度,计算公式如下:

$$P_k = \sum \theta_{ki} / N \dots \dots \dots (5)$$

其中,  $P_k$  表示第  $k$  个主题的主题强度,  $\theta_{ki}$  表示第  $i$  篇文档的主题是主题  $k$  的概率,  $N$  为政策文本的总数。本研究所得 8 个主题的主题强度分布见图 6。

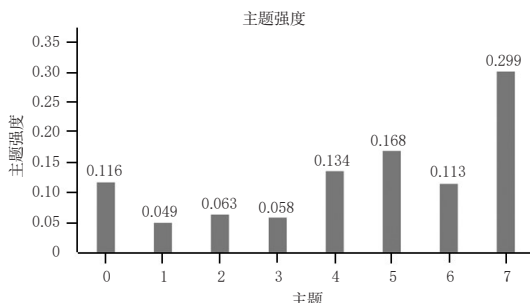


图6 8个主题的主题强度分布

Fig 6 Topic intensity distribution of 8 themes

参考文献的分类方法<sup>[17]</sup>,根据上述主题强度值对主题挖掘结果进行热门主题和冷门主题的对比研究。强度较高的主题有 7、5、4,即为热门主题;主题 0、6 强度适中,主题 1、2、3 强度明显低于其他主题,即为冷门主题。

4.2.2 主题内容分析 由“2.2”项下模型和本课题组对政策文本的先验知识,结合主题的代表特征词进行归纳总结,可得主题-特征词分布,详见表 2(因篇幅所限,本文仅列出每个主题所对应的前 10 个特征词)。

根据各个主题-特征词的分布情况,对 8 个主题的侧重点和政策关注的主要层面归纳如下:主题 0 为项目金融支持,其特征词“项目”“投资”“外商投资”表明政策对项目推进的支持侧重于资金的投入;主题 1 为中小企业建设,“新兴产业”“科技攻关”“优惠政策”体现了政策对于中小企业科技创新的支持;主题 2 为资源引领,“上市公司”“跨国公司”等词汇体现出政策对龙头企业的引领作用;主题 3 为出口贸易,“出口”“重大项目”“专利技术”所体现的是政策对于通过扩大海外市场提升竞争力的支持;主题 4 围绕的是融资支持,“融资”“资金”“扶持”等词表明政策的导向是通过促进融资来支持企业的研发创新;主题 5 与人才建设有关,“资助”“经费”等表明政策支持高层次人才、专家学者来共同参与建设国家高新区的生物医药产业;主题 6 是税收优惠,“所得税”“抵扣”等说明政策在税收方面实行优惠以促进生物医药企业的发展;主题 7 为技术创新,“科技”“创新”“研发”等表明创新是生物医药产业发展的动力源泉,是政策支持的重点对象。结合图 6 可知,国家高新区 518 项生物医药产业政策中的强度较高的热门主题依次为技术创新(0.299)、人才建设(0.168)、融资支持(0.134),强度较低的冷门主

表2 主题-特征词分布

Tab 2 Topic-feature word distribution

主题	特征词	权重	主题	特征词	权重		
主题0项目金融支持	税收	0.069 7	主题4融资支持	上市	0.080 4		
	固定资产	0.062 6		融资	0.048 9		
	项目	0.051 4		资金	0.029 7		
	土地	0.049 7		扶持	0.028 5		
	投资	0.034 0		基金	0.028 2		
	外商投资	0.030 9		投资	0.022 9		
	缴纳	0.029 2		注册资本	0.022 0		
	财政	0.028 3		补助	0.020 5		
	投资额	0.027 4		减免	0.020 2		
	设备	0.025 1		股权	0.020 1		
	主题1中小企业建设	新兴产业		0.143 4	主题5人才建设	人才	0.109 6
		名牌产品		0.124 0		高层次	0.042 0
		留学人员		0.095 6		资助	0.036 3
战略性		0.090 0	补贴	0.034 0			
创业园		0.058 8	团队	0.030 8			
科技攻关		0.024 2	经费	0.030 3			
符合条件		0.016 0	专家	0.028 9			
优惠政策		0.006 3	创业	0.028 6			
中型企业		0.001 5	引进	0.027 4			
产业集聚		0.001 5	科研	0.023 9			
主题2资源引领		房租	0.056 6	主题6税收优惠		所得税	0.061 2
		上市公司	0.056 4			纳税	0.052 8
		跨国公司	0.052 9			高新技术	0.046 5
	用房	0.046 7	抵扣		0.041 0		
	利率	0.046 5	经济		0.034 8		
	专项基金	0.044 2	所得额		0.033 3		
	专业化	0.043 2	征收		0.027 3		
	增资	0.040 3	增值税		0.026 7		
	房产	0.034 0	销售		0.025 3		
	银行	0.038 6	新产品		0.023 1		
	主题3出口贸易	出口	0.136 9		主题7技术创新	补助	0.026 7
		退税	0.081 0			创业	0.022 4
		重大项目	0.022 1			科技	0.021 9
免收		0.065 7	创新	0.021 7			
医药		0.057 3	平台	0.021 7			
投资者		0.053 5	资助	0.020 0			
药品		0.038 7	补贴	0.019 5			
房地产		0.036 1	研发	0.018 4			
专利技术		0.029 4	技术	0.018 1			
出资		0.027 7	支持	0.018 1			

题为资源引领(0.063)、出口贸易(0.058)、中小企业建设(0.049),强度中等的主题为项目金融支持(0.116)、税收优惠(0.113)。

## 5 分析与讨论

本研究在构建LDA主题模型的基础上,对2010年1月—2019年12月涉及生物医药产业的518件国家高新区政策文本进行了量化分析,辅助以词频分析等方法,较为清晰地分析了政策的总体分布和主题的强弱,避免了人工解读政策文本的主观性,证实了LDA主题模型用于政策文本量化分析的可行性,为松散、大量的政策文本的解读提供了有效途径,为科学、详尽地了解各园区生物医药产业政策导向提供了有价值的信息。由政策主题强度分布及主题内容分析可知,国家高新区生物

医药产业政策体系的建设存在以下不足。

### 5.1 生物医药创新激励措施运用失衡

由上述分析可见,我国国家高新区促进生物医药产业发展的激励措施存在较明显的偏向性,技术创新、人才建设、融资支持这3个热门主题的程度达0.601,由热门主题中的特征词“融资”“补助”“资助”结合政策文本高频词中的“补贴”(频次1542,占比2.63%)、“资金”(频次1443,占比2.46%)、“补助”(频次1295,占比2.21%)可见,政策多以资金补贴为主要激励措施。生物医药产业创新发展需要政府强有力的政策激励措施来支持,但激励措施运用失衡同样会对产业发展产生不利影响<sup>[21]</sup>。如果政府研发补贴等直接资金支持投入过重,容易导致企业忽视自身产业发展规律,盲目扩张、产能过剩;而适当的税收优惠政策可以通过降低企业的创新成本,引导企业加大研发投入,进而享受更多税收优惠额,有利于企业根据自主创新需求有针对性地扩大投入,形成长期的激励效果<sup>[22]</sup>。目前,我国高新区生物医药产业的税收优惠政策和体系仍然有待完善,主要存在优惠范围窄、力度小、人力资本和组织运营投入相关税收政策缺失等问题<sup>[23]</sup>,故需要进一步健全创新激励导向的税收制度体系,并还应在运行过程中不断优化系统规范和实施细则。总之,鼓励生物医药企业创新,需要综合运用各项相辅相成的配套激励措施,不断提高政策运用的科学性、可操作性,方能达到最优的效果。

### 5.2 对企业扩大海外市场缺乏足够的支持力度

资源引领、出口贸易为冷门主题,主题强度分别为0.063、0.058,是整个国家高新区生物医药产业政策体系中建设较为薄弱的部分。生物医药产业作为我国重点发展的战略性新兴产业,已成为多个国家高新区的支柱产业,正处于蓬勃发展阶段。但在全球医药市场中,仍是起步较早的欧美国家在技术和产品上占据领先地位<sup>[24]</sup>。技术和产品的领先重在技术创新和产品质量,通过以上主题分析结果可以得知,我国现有政策主要通过研发补贴、人才引进等举措来促进企业的技术创新,缺少对企业引进海外先进技术和设备的支持,也缺少支持企业间、企业与国内外高校和研究机构间交流合作的配套政策,较难发挥行业龙头的带动作用。产品、技术和贸易是我国生物医药产业占据国际领先地位的必要条件,而目前我国生物医药产业国际竞争力还需进一步提升<sup>[25]</sup>,针对企业知名品牌打造、业务国际化发展的政策也处于相对匮乏的状态,亟待相关部门出台相关政策给予指导。

### 5.3 对中小企业扶持力度不足,缺少具体的针对性政策

中小企业建设的主题强度为0.049,是8个主题中强度最低的冷门主题,可见各国家高新区对于中小生物医药企业的扶持还欠缺针对性的政策和措施。从主题内

容来看,生物医药产业政策涉及中小企业建设的主要为企业的创新能力和其中“名牌产品”“科技攻关”以及“优惠政策”等正是我国中小生物医药企业所面临的且亟待解决的问题。例如,中小生物医药企业作为新药创新的主体主要会面临技术、创新人才以及资金不足等困难;组织化程度低的中小生物医药企业还缺少合适的融资平台和渠道,很难获得融资;此外,中小生物医药企业税费负担重也一定程度上阻碍了其长远发展<sup>[26]</sup>。2014年,国务院就出台了《关于扶持小型微型企业健康发展的意见》<sup>[27]</sup>来引导各级政府、各地高新区构建中小企业良好的发展和创新环境,加强对中小企业的扶持力度,但从目前国家高新区政策分析结果来看,对于中小生物医药企业创新、税收以及融资等方面仍缺乏详细的指导政策和配套措施,有待进一步完善。

## 6 结语

综上所述,目前我国国家高新区生物医药产业政策以技术创新、人才建设、融资支持为主,缺乏对中小企业建设、资源引领、出口贸易等的关注。今后国家高新区对生物医药产业的政策支持及体系建设应重点完善政策激励措施,平衡各领域政策运用比例;鼓励企业扩大海外市场,学习海外先进技术,加强科研等领域的多边合作;配套相应的政策以增进企业间的交流合作,发挥龙头企业和名牌产品的带动作用,促进中小企业的发展。

## 参考文献

[1] 王胜光,朱常海.中国国家高新区的30年建设与新时代发展:纪念国家高新区建设30周年[J].中国科学院院刊,2018,33(7):693-706.

[2] 陈露,褚淑贞.我国生物医药产业工业园区现状及发展模式[J].现代商贸工业,2016,37(9):3-5.

[3] 程凌华,李享.中国火炬统计年鉴:第1部分:国家高新技术产业开发区[M].北京:中国统计出版社,2018:30.

[4] 王树海,闫耀民.中国火炬统计年鉴:第1部分:国家高新技术产业开发区[M].北京:中国统计出版社,2008:17.

[5] 卫平,张跃东,姚潇颖.国内外科技园区发展模式异质性研究[J].中国科技论坛,2018(7):180-188.

[6] 胡德巧.政府主导还是市场主导:硅谷与筑波成败启示录[J].中国统计,2001(6):16-18.

[7] 郑代良,钟书华.中国高新技术政策30年:政策文本分析的视角[J].科技进步与对策,2010,27(4):90-93.

[8] 杨慧,杨建林.融合LDA模型的政策文本量化分析:基于国际气候领域的实证[J].现代情报,2016,36(5):71-81.

[9] 赵乐,张兴旺.面向LDA主题模型的文本分类研究进展与趋势[J].计算机系统应用,2018,27(8):10-18.

[10] 郎政.大数据视野下中央与地方政府职能演变中的匹配度研究:基于甘肃省14市(州)政策文本主题模型(LDA)[J].情报杂志,2018,37(9):78-85.

[11] 刘江华.1种基于Kmeans聚类算法和LDA主题模型的

文本检索方法及有效性验证[J].情报科学,2017,35(2):16-21,26.

[12] BLEI DM, NG AY, JORDAN MI. Latent Dirichlet allocation[J]. *J Mach Learn Res*, 2003, 3(4/5):993-1022.

[13] GRIFFITHS TL, STEYVERS M. Finding scientific topics[J]. *Proc Natl Acad Sci: USA*, 2004, 101(1):5228-5235.

[14] HANCHEN J, MAOSHAN Q, PENG L. Finding academic concerns of the Three Gorges Project based on a topic modeling approach[J]. *Ecol Indic*, 2016. DOI: 10.1016/j.ecolind.2015.08.007.

[15] ROBERTSON S. Understanding inverse document frequency: on theoretical arguments for IDF[J]. *J Doc*, 2004, 60(5):503-520.

[16] BOTH A, HINNEBURG A. *Exploring the space of topic coherence measures[C]//Eighth ACM international conference on web search and datamining*. New York, Association for Computing Machinery, 2015:399-408.

[17] 李湘东,张娇,袁满.基于LDA模型的科技期刊主题演化研究[J].情报杂志,2014,33(7):115-121.

[18] 中关村管委会.中关村国家自主创新示范区天使投资和创业投资支持资金管理暂行办法[EB/OL].(2014-11-01)[2020-10-27]. <http://zgcgw.beijing.gov.cn/zgc/zwgk/zcfg18/sfq/139714/index.html>.

[19] 泰州医药高新区管委会.关于加强中国医药城专业化高层次人才队伍建设的实施意见[EB/OL].(2019-01-02)[2020-10-27]. [http://gxqxxgk.taizhou.gov.cn/art/2019/1/2/art\\_32431\\_3204.html](http://gxqxxgk.taizhou.gov.cn/art/2019/1/2/art_32431_3204.html).

[20] 泰州医药高新区管委会.关于推进科技创新引领高质量发展的若干政策[EB/OL].(2019-02-21)[2020-10-27]. [http://gxqxxgk.taizhou.gov.cn/art/2019/2/21/art\\_32435\\_3264.html](http://gxqxxgk.taizhou.gov.cn/art/2019/2/21/art_32435_3264.html).

[21] 毕晓方,张俊民,李海英.产业政策、管理者过度自信与企业流动性风险[J].会计研究,2015(3):57-63,95.

[22] 曾萍,邬绮虹.政府支持与企业创新:研究述评与未来展望[J].研究与发展管理,2014,26(2):98-109.

[23] 寇恩惠,李向阳,赵杨.完善创新型小微企业税收优惠政策的思考[J].财政监督,2017(14):5-10.

[24] 刘泉红,刘方.中国医药产业发展及产业政策现状、问题与政策建议[J].经济研究参考,2014(32):39-67.

[25] 潘琪,王隽,徐晓媛,等.“一带一路”背景下我国医药产品进出口现状与发展[J].中国医药导报,2019,16(5):168-171.

[26] 孙振淋,沈念伍,褚淑贞.医药产业集聚中的中小医药企业[J].中国药事,2013,27(3):288-291,297.

[27] 国务院.关于扶持小型微型企业健康发展的意见[EB/OL].(2014-11-20)[2020-10-27]. [http://www.gov.cn/zhengce/content/2014-11/20/content\\_9228.htm](http://www.gov.cn/zhengce/content/2014-11/20/content_9228.htm).

(收稿日期:2020-02-09 修回日期:2020-10-28)

(编辑:孙冰)