

“强度-内容”视域下我国公共数字文化政策主题演化研究

侯艳辉, 赵文静, 王家坤

(山东科技大学 经济管理学院, 山东 青岛 266590)

摘要:为分析我国公共数字文化政策主题分布及其演化路径,有效揭示政策主题变迁趋势,把握政策主题演化深层机制,为政策制定与应用提供理论支持。收集我国公共数字文化政策文本,运用 LDA 模型识别政策文本主题及其关键词;构建融合多特征的主题强度表征模型,分析主题强度演化情况;利用 Word2Vec 模型获取主题关键词最相似词集合,分析主题内容演化情况。研究发现,我国公共数字文化政策主要有三大惠民工程建设、均等化配置、建设与保障管理和数字文化产业发展四大主题,主题演化呈现阶段性变化、科学技术驱动、聚焦发展问题等特点。基于主题演化分析结果,未来公共数字文化建设应紧密围绕国家战略导向,加强科技创新与人才培养;积极响应公众数字文化需求,促进共建共享合作模式;聚焦发展问题与矛盾,缩小不同区域与群体间发展差距。

关键词:公共数字文化;主题强度;多特征融合;主题内容;主题演化

中图分类号:G353.1

文献标识码:A

文章编号:1008-7699(2025)04-0081-12

一、引言

“公共数字文化”是由“公共文化”和“数字文化”有机结合形成的复合概念,是互联网环境下公共文化向网络化、信息化和数字化的延伸和拓展。公共数字文化以数字技术为依托,具有辐射面广、传播速度快、资源广泛共享等特点。推动公共数字文化建设对有效满足人民群众精神文化生活需要,提高文化服务普惠化、公平性,以及实现文化资源的数字化传播及保护,提高国家文化软实力具有重要意义。

当前,我国公共数字文化建设已取得显著成绩,但仍面临着一些亟待解决的问题,例如,数字文化资源整合效果欠佳、资源共享与知识产权保护冲突等,这些问题的存在制约着我国公共数字文化的进一步发展。公共数字文化政策作为公共数字文化建设的基石,是促进公共数字文化长效发展的必要保障。通过分析公共数字文化政策主题演化过程,可以深入了解政策关注重点变化情况和政策内容更迭情况,探究政策变迁深层机制,对于解决现存问题,为后续政策制定与应用提供支持和引导,进一步推进公共数字文化建设。因此,本文收集 2011—2022 年间我国公共数字文化相关政策文本,利用隐含狄利克雷分布(Latent Dirichlet Allocation, LDA)模型识别政策主题,从政策主题强度和主题内容两方面综合分析主题演化过程,揭示政策主题发展脉络及规律,以期提高公共数字文化领域政策制定与实施的科学性和合理性,为解决当前发展中存在的问题提供理论支持。

二、相关研究

政策主题演化是指在识别政策文本潜在主题结构的基础上,探究主题时序发展变化情况,目前主要

收稿日期:2024-09-24

基金项目:山东省社会科学规划研究项目(22DGLJ27)

作者简介:侯艳辉(1978—),男,山东潍坊人,山东科技大学经济管理学院副教授、硕士生导师,博士。

应用的定量分析方法大致可分为两类。一类是文献计量学方法,通常以内容分析法为基础,通过人工编码文本主题,借助频次分析、共词网络分析等手段对文本主题进行量化分析,以挖掘政策主题特征及其演化规律。李放等^[1]构建了主题、工具、属性三维框架,基于内容分析法对家庭养老支持政策文本内容进行了人工编码,结合频数统计,从单维度和交叉维度实证分析了政策主题情况;王旭等^[2]¹²⁶从工具、效力、主题三个维度对个人信息保护政策文本内容进行了人工编码,结合主题词的语义网络分析了政策主题内容及其相关联系。另一类为文本挖掘类方法,利用文本聚类、主题模型等对政策文本进行结构化处理和量化分析,从中提取主题信息并以此为基础展开演化研究。吴鹏等^[3]借助 LDA 主题模型识别政策主题,分析了时间维度和层级维度上新冠疫情政策主题演变情况;Ma 等^[4]利用 LDA 模型识别政策主题,分析了家庭教育政策的时空演化格局。更进一步,胡吉明等^[5]融合词嵌入模型 Word2Vec 和 LDA 模型,结合政策文本上下文语义识别政策主题,分析了“互联网+”政策文本的主题特征。

随着大规模数据处理需求的持续升级和自然语言处理技术的快速发展,越来越多的学者利用文本挖掘类方法分析政策主题,以期更加客观、快速地解读政策内容。但目前关于公共数字文化政策的研究多采取质性分析方法或基于政策工具、政策主体等外部属性展开,针对政策主题的研究相对较少,且多以文献计量学方法为主。Wright 和 Gray^[6]采用质性分析方法对政策文本进行解读,分析了文化与数字之间的关系及英国国家内部文化政策状况;吴丹等^[7]基于融合创新视角,详细解读了我国公共数字文化政策发展历程。陈庚等^[8]以主题词分析法和语义网络分析法为主,量化分析了我国不同阶段数字文化产业政策主题特征,进一步揭示了主题演化过程;朱益平等^[9]采用文献计量法提取了政策文本高频主题词,通过共词聚类识别政策主题,对我国公共数字文化服务政策主题变迁进行了系统分析。

在识别文本主题的基础上,主题演化研究主要从主题强度演化研究和主题内容演化研究两方面展开。^[10,11]主题内容反映了主要内涵与核心观点,主题强度反映了关注度与影响力,两者的结合能够更全面地展示一个主题的内在特性与关注价值。主题强度又称为主题热度或主题关注度,它能够反映某时间段一个主题的重要性或显著程度,从而识别出领域内值得关注的主题和热点问题。主题强度表征方式主要有以下两类。一是利用文本内部特征即主题概率分布表征主题强度,如主题平均强度值、主题文档支持率等。这类计算指标较为传统且常见,郭强等^[12]采用主题平均强度值反映了农业农村科技创新政策的主题热度与关注度,Wu 等^[13]借助主题平均强度揭示了建筑废弃物研究领域的主题热点和发展趋势。二是利用文本外部特征,如主题文档数量比、主题篇均被引量、期刊影响因子权重等表征主题强度。这类指标常见于学术文献研究,刘自强等^[14]⁹²使用主题篇均被引量对碳纳米管研究领域的相关文献进行了主题演化分析,霍朝光等^[15]³⁵结合期刊影响因子和主题出现频次构建主题强度表征指标,分析了图书馆与信息科学(Library and Information Science, LIS)领域国内外核心期刊数据主题热度情况。

主题内容演化主要采用语义相似度计算或主题共现关联等方法,分析主题语义细节变化和整体主题变迁情况,王晰巍等^[16]利用 JS 散度反映了网络舆情传播中社会群体主题演化情况,Wang 等^[17]借助主题共现频次分析了生物创新领域主题内容演化,王伟等^[18]借助关键词间的余弦相似度分析了众筹项目在线评论主题的内容演化。随着神经网络模型的兴起,学者开始将 Word2Vec 模型、BERT 模型等应用到文本主题分析,进一步提高了关联关系计算准确度。周爱霞等^[19]³⁷利用 Word2Vec 模型及余弦相似度获取了“大数据”最相似词语,通过比较最相似词语对大数据关注热点进行了分析;胡泽文等^[20]结合 Word2Vec 模型、BERT 模型及余弦相似度分析了颠覆性技术领域主题的相互演化规律。

综上所述,目前我国公共数字文化政策主题演化定量研究成果相对较少,且存在以下问题:(1)研究方法多采用内容分析法、人工编码形式,缺乏对自动化文本处理技术的应用,存在专家判断主观性较强的缺陷;(2)政策文本主题强度测度多依据单一的文本内部语义特征,忽略了政策发文类型、政策关联情况等的影响,存在一定局限性。鉴于此,本文采用 LDA 模型,通过无监督学习识别我国公共数字文化政策潜在主题及其关键词,构建融合多特征的政策主题强度表征模型,分析文本主题强度演化过程,并引入

Word2Vec 模型对比热点主题关键词最相似词情况,分析文本主题演化过程。从主题强度和-content 两方面系统揭示政策文本主题发展整体脉络与规律,以期提高公共数字文化领域政策制定与实施的科学性,为解决当前发展问题提供理论参考。

三、研究设计

本文研究过程分为三个阶段:第一阶段为数据处理与主题识别,收集我国公共数字文化相关政策文本,利用 LDA 主题模型获取政策文本主题—关键词及文档—主题概率分布;第二阶段为主题强度演化研究,分别采用主、客观赋权法对政策文本语义特征、关联特征和效力特征进行赋权,利用博弈论方法求解组合权重,构建融合多特征的综合指标分析主题强度演化过程;第三阶段为主题内容演化研究,引入 Word2Vec 词嵌入模型,对比不同阶段各主题对应关键词最相似词的变化情况,分析主题内容演化过程。本文研究框架如图 1 所示。

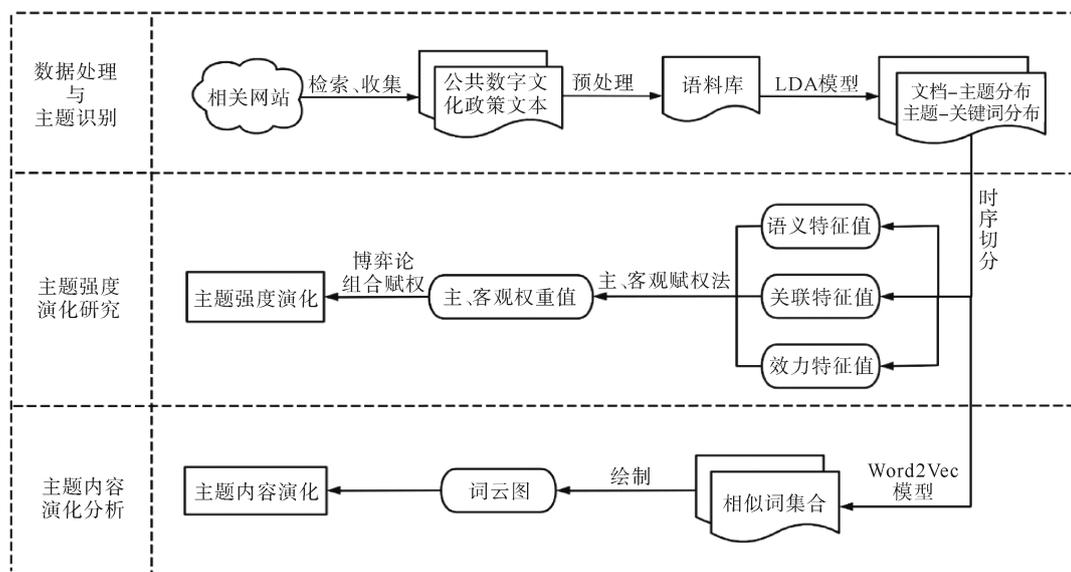


图 1 研究框架图

(一)主题识别

LDA 主题模型作为一种自动化文本处理技术,能够识别文档集隐含的主题结构,自动生成“文档—主题”概率分布和“主题—关键词”概率分布。^[21]模型具有良好的扩展性,且学者们根据自身研究需要,以 LDA 为基础进行了一系列联合应用,进一步增强了文本处理的效率与效果,该模型现已成为主题识别领域应用最广泛的模型之一。^[22,23]鉴于此,本文选取 LDA 模型识别政策文本潜在主题,通常情况下,LDA 模型主题数 K 需要预先设定。困惑度(Perplexity)是衡量 LDA 模型预测能力的标准方法,其值越低,表明模型预测能力越强,有效性越高。^[24]本文基于困惑度确定主题数,通过绘制主题数—困惑度折线图,观察困惑度变化趋势,选取最优主题数。

(二)基于多特征融合的主题强度表征模型

1. 主题强度多特征表示

为避免先离散时间方法造成主题不对齐问题,本文采用后离散时间方法,先完成整个数据集的主题识别,然后将数据划分至若干时间段,进而表征各时间段的政策主题强度。政策文本结构可以分成语义特征、内容特征和外部特征三个部分,继而从不同角度反映政策主题的关注程度与影响力。^[25]具体而言,

文本关键词、短语等语义特征能够通过词频等形式揭示政策文本包含的重点,从政策文本的核心问题与目标反映政策主题强度;引用与被引用关系等内容关联特征能够展示政策文本的影响力,从政策文本的信息扩散程度与逻辑结构角度反映政策主题强度;发文主体、发文类型等外部结构特征所具有的法律效力与执行力度,能够从政策文本的效力程度与权威性角度反映政策主题强度。基于此,本文依据政策文本结构,融合语义特征、关联特征和效力特征,构建主题强度多特征表征模型,全面评估政策主题的主题强度与影响力。

(1)语义特征

目前,关于政策文本主题强度的研究绝大部分依据文本语义特征即主题概率分布,从文本潜在含义层面表征主题强度,主要采取三种计算方式:基于主题支持的文档数量、基于语料库的话题概率、基于文本话题的主题显著性^[26]。本文参考曾江峰等^[27]的研究,选取目前研究成果中应用最广泛的主题平均强度指标描述政策主题语义特征。计算方法如式(1)所示:

$$S_{t,k} = \frac{\sum_{d=1}^{D_t} \theta_{d,k}}{d_t}, \tag{1}$$

其中: $S_{t,k}$ 为时间段 t 上主题 k 的平均语义强度, $\theta_{d,k}$ 为第 d 篇政策文本第 k 个主题的概率, D_t 为时间段 t 上政策文本数量。

(2)关联特征

主题对应文本被引量,可从文本内容关联作用层面反映随时间推移,主题受关注程度的变化。因此,本文借鉴学术文献主题强度研究中的主题篇均被引量,^{[14]93} 利用主题对应政策文本引用情况描述政策主题关联特征。主题对应政策文本被引用次数越多,说明该文本对应主题受关注程度越高,影响力越大。计算方法如式(2)所示:

$$AC_{t,k} = \frac{SumC_{t,k}}{SumD_{t,k}}, \tag{2}$$

其中, $AC_{t,k}$ 为时间段 t 上主题 k 的篇均被引量, $SumC_{t,k}$ 为时间段 t 上主题 k 对应政策文本被引总量, $SumD_{t,k}$ 为时间段 t 上主题 k 对应政策文本总量。

(3)效力特征

学术文献主题强度相关研究利用期刊影响因子加权,考量发布在不同期刊的文献所产生的学科影响力差异。^{[15]39} 同理,政策文本由不同主体、以不同类型颁布,产生的影响力不同。因此,本文借鉴相关学者对政策文本效力划分与赋值的标准,依据政策发文机关层级及政策类型分别对文本进行赋值,并将二者加和,得到表征政策文本主题的效力值,从政策外部属性层面反映主题强度^{[2]128[28]}。政策效力赋值标准见表 1,效力特征值计算方法如式(3)所示:

$$V_{t,k} = \frac{SumTA_{t,k}}{SumT_{t,k}}, \tag{3}$$

其中, $V_{t,k}$ 为时间段 t 上主题 k 的平均效力值, $SumTA_{t,k}$ 和 $SumT_{t,k}$ 分别为时间段 t 上主题 k 对应政策文本效力赋值总和与文本数量总和。

表 1 政策效力赋值表

政策主体	赋值	政策类型	赋值
全国人民代表大会及其常务委员会	5	法律法规及纲要	4
中共中央办公厅、国务院办公厅	4	条例、规定、规划等	3
国家各部委	3	意见、办法、方案等	2
国家各部委各部门、省级人民代表大会及其常务委员会	2	通知	1
省级政府部门	1		

关于关联特征和效力特征中主题对应文档问题,本文依据 LDA 主题模型结果,选取平均概率值作为阈值,即当一篇文档中超过设定阈值的词是由特定主题生成时,可将该文档看作该主题支持文档。

2. 主题强度多特征融合

本文借助权重系数实现上述多特征融合。赋权方法分为主观赋权法和客观赋权法,主观赋权法主要依据决策者经验或偏好赋予指标适当权重,客观赋权法则根据实际数据反映的客观信息确定指标权重。从方法的优劣势来看,主观赋权法汇聚了专家的专业知识与经验,但忽略了客观数据的支持;客观赋权法侧重于客观数据信息,却忽略了专家经验和见解。由此,为得到科学合理的赋权结果,应对两者进行集成和融合。^[29]在多种组合赋权理论和方法中,博弈论组合赋权法能够考虑多方利益和策略选择的影响,优化多因素影响下的目标,寻求兼顾各方的最优解,是应用最广泛的组合赋权方法之一。因此,本文首先分别使用主观赋权法中的层次分析法和客观赋权法中的熵权法,对三个特征进行赋权。然后,利用博弈论组合赋权法综合主、客观赋权结果,在主、客观赋权法确定的权重中寻找一致或妥协,即确定一个最优组合权重,提高各特征权重赋值的均衡性和科学性。^[30,31]最后,通过最优组合权重确定各主题在各个时间段内的综合强度,并借助热力图进行可视化,从而分析主题强度演化过程。

假设主、客观赋权法获得的权重转置向量分别为 w_1^T, w_2^T , 利用式(4)将各权重矩阵进行线性组合优化得最小优化值

$$\min \left\| \sum_{j=1}^2 \alpha_j w_j^T - w_i^T \right\|_2, (i = 1, 2; j = 1, 2)。 \quad (4)$$

依据矩阵微分性质,可将式(4)转化为最优化一阶导数条件的线性方程组:

$$\begin{pmatrix} w_1 w_1^T & w_1 w_2^T \\ w_2 w_1^T & w_2 w_2^T \end{pmatrix} \begin{pmatrix} \alpha_1 \\ \alpha_2 \end{pmatrix} = \begin{pmatrix} w_1 w_1^T \\ w_2 w_2^T \end{pmatrix}。 \quad (5)$$

利用式(5)求出权重最优分配系数 (α_1, α_2) , 得到权重最优分配系数 α^* :

$$\alpha_i^* = \frac{\alpha_i}{\sum_{k=1}^2 \alpha_k}, i = 1, 2。 \quad (6)$$

最终求得最优组合权重矩阵 w^* :

$$w^* = \alpha_1^* w_1^T + \alpha_2^* w_2^T。 \quad (7)$$

(三)基于 Word2Vec 的主题内容演化分析

Word2Vec 模型是一种广泛应用于自然语言处理领域的词向量表示模型,其通过将词语映射为连续向量空间中的实数向量,捕捉与表示词语间的语义关系。该模型有两种主要实现算法:CBOW (Continuous Bag of Words)和 Skip-gram,^[32]其中, Skip-gram 模型通过当前目标词预测上下文词。本文参考周爱霞等^{[19]39}提出的最相似词比较方法,用 Skip-gram 模型对预处理后的各阶段语料库训练 Word2Vec 模型;再结合 Word2Vec 模型得到的词语词向量化表示,计算各阶段语料库中所有词语与各主题关键词的余弦相似度,获取各时间段中各主题对应关键词的最相似词集合,并依据最相似词词频绘制词云图,比较分析主题内容演化情况。

四、公共数字文化政策主题演化

(一)数据收集与处理

2011年11月15日文化部和财政部出台《关于进一步加强公共数字文化建设的指导意见》,正式提出“公共数字文化建设”概念,在此之前,“公共文化”和“数字”并未融合成新概念。^[33]基于此,本文检索范围定为该政策出台之日至2022年12月31日。在北大法宝、中国政府网、中国文化和旅游部等相关网站,以“数字文化”作为关键词进行全文检索。由于各地区数字文化基础不同,发展要求各有差异;且地方性

政策覆盖范围有限,通常仅适用于特定地区;因此本文仅收集中央和省级政府部门所颁布的政策文本,对其他地方政策法规不予选取。同时,剔除批复、函、工作报告等非正式文件以及政策内容与数字文化关联度不高的文本,共收集 110 份政策文本,年度发文量如图 2 所示。对所获文本进行预处理,首先选取有关数字文化内容的政策文本全文或章节;其次根据数据集实际内容构建自定义词典和停用词表,完成词性标注等工作,并利用 jieba 分词包对文本进行分词得致政策文本语料库。

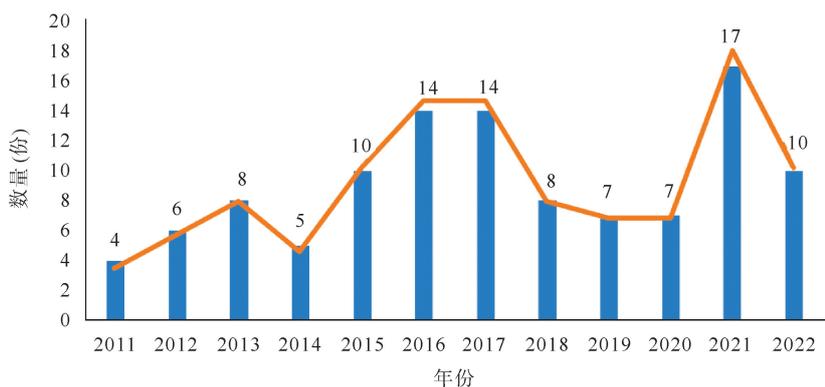


图 2 我国公共数字文化年度发文量

(二)主题识别结果

1. 基于困惑度的主题数确定

通过主题数-困惑度折线图确定最优主题数,一般情况下,当困惑度最小或处于拐点处时,模型效果最佳。图 3 为本文所绘制的主题数-困惑度折线图,可以看出,随着主题数(Number of Topic)增大,困惑度(Perplexity)逐渐减小,主题数为 4 时,出现拐点且困惑度达到最低点。因此,本文最终选定主题数为 4。

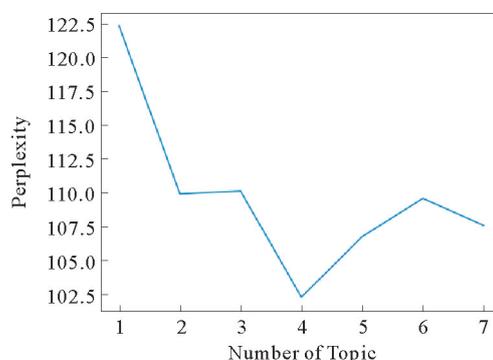


图 3 主题数-困惑度折线图

2. 主题识别与分析

确定主题数后,使用 LDA 模型对预处理后的整个数据集进行主题识别,得到主题-关键词概率分布及文档-主题概率分布,主题识别结果见表 2。其中,依据主题-关键词概率分布,概率值从大到小依次列举出与各主题相关的 10 个关键词,并联系主要关键词标识主题。

表 2 主题-关键词概率分布表

主题	主题名称	关键词
Topic 1	三大惠民工程建设	文化共享工程、基层、电子阅览室、数字图书馆、机构、数字文化资源、资源建设、服务网络、公共文化服务体系、重点
Topic 2	均等化配置	贫困地区、县域、基层综合性文化服务中心、农牧民、边疆万里数字文化长廊、服务点、未成年人、传统媒体、少数民族语言数字资源、数字文化驿站
Topic 3	建设与保障管理	设施管理单位、数据资料、网络服务、主管部门、卫星网络、文化信息资源库、群众文艺、文化行政部门、宽带互联网、办法
Topic 4	数字文化产业发展	旅游、生产、消费、演艺、产品、数字创意、融合、文旅、人工智能、行业

(三)主题强度演化分析

1. 主题强度多特征测度

(1)多特征计算结果

完成整个数据集的主题识别后,综合考虑我国公共数字文化政策发文数量变化趋势及数字文化相关研究,^[34-36]将数据集划分为4个时间段:第一阶段为2011—2014年,第二阶段为2015—2016年,第三阶段为2017—2019年,第四阶段为2020—2022年。按照公式(1)、(2)、(3),依据LDA模型所获文档—主题概率分布计算语义特征;基于北大法宝数据库中的“引用本法”计算政策文本间的引用关联特征;依据政策效力赋值表计算效力特征;得到各时间段主题强度三个特征值。为避免特征值间存在类型、量纲不一致等问题影响研究结果,采用Min-Max归一法对原始特征值进行无量纲化处理,得到主题强度多特征表示结果。

(2)主、客观赋权结果

采用层次分析法确定特征主观权重,设计政策文本主题强度特征权重调查问卷,选择公共政策研究领域6位专家发放问卷,收集专家判断矩阵数据。6位专家中,3位为省、市级文化和旅游部门管理人员,3位为高校公共政策研究领域专家学者。经计算,6位专家的判断矩阵均满足一致性检验。依据调查问卷结果,采用算术平均值计算公共数字文化政策主题强度语义特征、关联特征和效力特征主观权重,分别为0.6348、0.1672、0.1980^①。

采用熵权法获取特征客观权重,为避免0值导致对数运算无意义,首先对归一化后的数据非负平移0.01;^[37]然后借助SPSS分析软件获得熵权法结果,具体见表3。

(3)博弈论组合赋权结果

依据上文博弈论组合赋权法及其求解思路,利用Matlab软件求解各主题主、客观权重最优分配系数,进而求得各主题三个特征最优综合权重,最后由综合权重得到各主题综合强度值,各综合权重系数见表4。

表3 熵权法结果汇总

主题	语义特征值	关联特征值	效力特征值
Topic 1	0.2961	0.4943	0.2096
Topic 2	0.2049	0.5285	0.2666
Topic 3	0.2370	0.5630	0.2000
Topic 4	0.2364	0.4936	0.2700

表4 博弈论组合赋权权重系数

主题	AHP-分配系数	熵权法-分配系数	语义特征-综合权重	关联特征-综合权重	效力特征-综合权重
Topic 1	0.6644	0.3356	0.5211	0.2770	0.2019
Topic 2	0.5781	0.4219	0.4534	0.3196	0.2270
Topic 3	0.5589	0.4411	0.4593	0.3418	0.1989
Topic 4	0.6280	0.3720	0.4866	0.2886	0.2248

2. 主题强度演化分析

根据上述过程获取的各主题综合强度值,结合时间段划分结果,构建公共数字文化政策主题强度变化时间序列,并绘制主题强度演化图,如图4所示。图中依据颜色深浅表示主题强度大小,主题强度越大,颜色越深;反之,颜色越浅,主题强度越小。

① 限于篇幅,不对具体过程进行展示,如有需要可联系作者获取。

分析图 4 可知,从整体演化趋势来看,公共数字文化领域四大政策主题强度变化都呈现一定波动现象。从各主题时序变化维度来看,Topic 1(三大惠民工程建设)和 Topic 4(数字文化产业发展)两个政策主题强度变化相对平稳,呈现出持续性存续的特点;与之相比,Topic 2(均等化配置)和 Topic 3(建设与保障管理)的强度变化较为剧烈,演化过程中出现强度骤升或骤降情况。结合具体政策内容分析可知,这与国家经济与社会发展规划的调整、科学技术发展以及建设过程中矛盾与问题的出现等有一定关系。

(四)主题内容演化分析

依据上述主题强度演化结果,由于 Topic 2 和 Topic 3 在个别时间段内主题强度偏低,且各阶段政策文本内容不存在相同高概率关键词,因此,本文仅对主题强度持续存续的 Topic 1 和 Topic 4 两个热点主题进行内容演化分析。根据研究设计所述步骤,获取了各时间段 Topic 1 和 Topic 4 对应概率前 5 个关键词(其可基本表明主题主要内容,且同时存在于各时间段政策文本中)的最相似词集合。实验表明,与各关键词最相似的前 10 个词语的余弦相似度相对较高,因此选取各关键词相似词集合的前 10 个词语,依据词频绘制词云图,分别如图 5、图 6 所示,可直观地揭示公共数字文化政策热点主题内容演化情况。

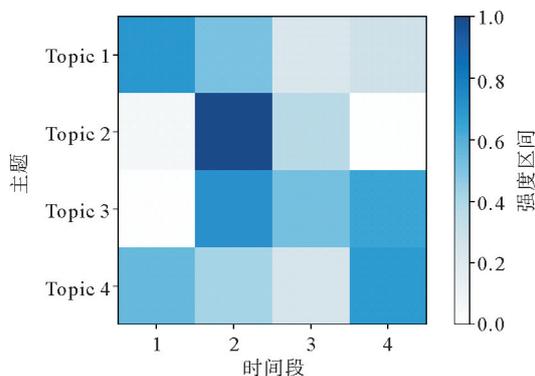


图 4 主题强度演化图



图 5 Topic 1 相似词词云图



图 6 Topic 4 相似词词云图

分析图 5 可知,2011—2014 年中,“基层、乡镇、社区、服务点、设施建设、数字资源库群、海量、节点”等词汇说明该阶段三大惠民工程建设主要以数字化资源建设、设施建设等为重点,为后续发展奠定基础。2015—2016 年中,“文化馆、图书馆、博物馆、文化信息资源、直播卫星、共建共享、宽带中国、服务点”等词

汇说明该阶段主要在前一阶段资源、设施建设基础上,以宽带中国、智慧城市等信息工程为依托,重点关注文化馆、图书馆等各类数字资源加工和服务平台建设。2017—2019年中,“文化馆、图书馆、文化共享工程、管理、一站式、数据分析、远程、界面”等词汇的出现说明该阶段进一步重视服务平台互联互通问题,推进集成化、一站式平台建设,促进公众供需有效对接;同时从“汉藏文化交流项目、残障人士”等词汇的出现可以看出,该阶段三大惠民工程建设与均等化配置内容相关联,进一步促进文化数字化全民共享。2020—2022年中,“服务、云展览、云视听、云阅读、智能服务、错峰”等词汇表明,该阶段依托新一代信息技术,公共文化云平台、云服务进一步发展,形成线上线下融合互动的公共数字文化体系。

由图6可知,2011—2014年中,“生产、产品、装备、消费、领域、文物、网络传播、数控系统”等词汇表明了该阶段数字文化产业发展主要立足于以文化装备产业为主的基础性产业发展,以推动优秀传统文化和当代精品网络文化相关产品生产、传播、消费的数字化进程。2015—2016年中,从“数字创意、内容、高新技术、产业链、创新链、贫困村、脱贫攻坚、边境”等词汇可以看出,该阶段一是重视以先进技术推动数字创意产业发展,促进产业链紧密衔接,转变产业发展格局;二是与Topic 2相关联,重视丰富贫困地区、民族地区等地区文化产品,以文化产业发展助力公共文化均衡发展。2017—2019年中,从“艺术品、形态、传统文化资源、商业模式、结构调整、传统文化产业、AR”等词汇可以看出,该阶段仍然以推动文化产业结构优化升级为重点,推进数字文化与传统文化、技术研发等相关产业融合发展。2020—2022年中,“新模式、多渠道、云旅游、内生、内容、传播”等词汇表明,该阶段仍以深化供给侧结构性改革为主线,推动产业链和创新链精准对接,以期不断优化产业结构;注重以科学技术为支撑,推动跨行业融合,其中尤其关注文化和旅游融合发展,构建和完善文旅内容生产新模式;同时从“个性化、差异化、定制消费”等词汇可以看出,该阶段力求通过统一平台,促进线上线下融合互动,进而精准把握公众需求,为公众提供个性化产品与服务。

五、总结与讨论

本文从“强度-内容”视域入手,借助LDA主题模型、融合多特征的主题强度表征模型及Word2Vec模型等,揭示了我国公共数字文化政策主题强度属性和内容属性的演化规律,为政策分析提供了综合的理论框架与方法体系,对我国相关部门推进公共数字文化建设具有一定应用价值。

(1)通过主题强度波动性变化的“均等化配置”“建设与保障管理”两个主题演化情况可知:公共数字文化政策的发展需要综合考虑社会、经济、技术等多维因素。在政策主题的演化过程中,这种发展主要体现为既受国家重大发展战略的引导,又由发展中的矛盾与问题所驱动。在均等化配置主题强度骤升阶段(2015—2016年),国家精准扶贫战略的实施以及数字文化建设过程中出现的数字鸿沟,进一步扩大问题,使得政府加大了对均等化配置相关政策内容的重视。如2015年中共中央办公厅、国务院办公厅印发《关于加快构建现代公共文化服务体系的意见》,明确提出按照精准扶贫要求,集中实施数字文化服务等一系列文化扶贫项目。随着脱贫攻坚战取得全面胜利,国家公共文化服务环境均等化、公平化建设取得一定成效,不同群体逐渐能够享受到更加公平、丰富的公共数字文化服务,均等化配置相关政策内容不再作为单独的重点内容出现,该主题强度减弱。

同时,为响应党的十八届三中全会关于构建现代公共文化服务体系的要求,加强公共数字文化建设与管理,促进公共数字文化科学规范发展,公共数字文化建设与保障管理相关内容受到重视。2015年文化部制定《文化部公共数字文化工程管理办法》,对公共数字文化工程建设相关部门职责设定、考核与奖惩机制等内容作出明确规定,不断完善工程建设组织管理,先后出台《文化部“十三五”时期公共数字文化建设规划》《公共数字文化工程融合创新发展实施方案》等政策,积极调整相关管理机制以破解发展难题、挖掘发展潜力。由此,建设与保障管理主题强度升高,并一直处于较高热度。

(2)通过主题强度持续性存续的三大惠民工程建设和数字文化产业发展两个主题演化情况可知:公

共数字文化政策主题内容初期主要集中于设施建设与资源整合等基础保障层面,后续在科学技术迅猛进步以及公众文化需求多元发展的推动下,政策主题不断丰富、完善,以实现让更大范围内的人民共享文化数字化成果的核心目标。三大惠民工程是实现文化数字化全民共享的根本条件和必要窗口,数字文化产业是文化产业发展的重要领域和数字经济的重要组成部分,两者始终是公共数字文化建设的主要内容,与之对应的两大政策主题始终在建设过程受到持续关注,主题强度变化相对平稳。

三大公共数字文化惠民工程,包括文化共享工程、数字图书馆推广工程和公共电子阅览室建设计划,是我国公共文化服务体系的重要基础性工程。发展前期,三大惠民工程建设相关主题内容主要以资源和设施等基础建设为重点,强调文化资源的数字化和共享。从宽带互联网的普及到无线网络的应用,再到云计算、虚拟现实的创新,科学技术的进步为三大惠民工程建设注入了强大动力。三大惠民工程建设相关政策内容开始更加重视依托科技发展不断促进数字文化资源整合与服务模式创新,并努力通过构建一站式、交互式服务平台与公众需求对接,以需求为导向推进公共数字文化发展。

随着经济的发展,民众收入水平和消费能力普遍提升,催发了多元化的数字文化消费需求;同时,科学技术的进步也为数字文化产品的生产与传播创造了全新机遇。相关政策内容顺应科技与市场发展趋势,关注重点从装备产业等传统基础产业转向充分利用云计算、虚拟现实、人工智能等新兴技术,以推进数字文化产业的多元化发展。通过颁布《关于推动数字文化产业高质量发展的意见》等政策,力求加快文化产业数字化布局,促进数字创意产业快速发展,为不同群体提供个性化、智能化、多样化的产品和服务。

基于以上对我国公共数字文化政策主题演化机制的展示,结合社会、经济、技术等发展现状,以及目前公共数字文化建设过程中存在的资源整合不足等问题,对未来公共数字文化领域发展提出以下建议。

一是,紧密围绕国家战略导向,通过科技创新与人才培养助推领域发展。公共数字文化政策主题演化方向始终与国家战略方向相一致,与社会、经济、技术发展相适应。因此,需牢牢把握国家创新驱动、人才强国等重大战略导向,通过科技创新与人才培养助推公共数字文化发展。在科技创新上,一方面,要依托信息技术构建统一的公共数字文化服务平台,为高效整合海量碎片化数字资源提供新路径;另一方面,要鼓励跨学科合作,促进科技与艺术融合,优化数字资源利用效能,助推数字文化产业升级。在专业人才上,要加大对数字文化专业人才的培养,建立多层次、多领域的教育培训体系,提升既懂技术又懂文化的复合型人才比例,为解决伴随文化数字资源开放与信息技术发展产生的文化数据安全监管、产权保护等问题提供支持。

二是,积极响应公众数字文化需求,促进共建共享合作模式。公众需求提升是公共数字文化政策演化的重要驱动因素,充分考虑并满足公众的数字文化需求与偏好是确保政策针对性与有效性的重要前提。因此,要健全文化交流平台和反馈机制,及时了解公众多样化的数字文化需求,为文化数字化产品的创作与传播提供依据,进而调节数字文化产业供需结构,提高数字文化服务效能,推动产业可持续性发展。同时,要鼓励、支持文化机构与社区、学校、企业等多方建立合作关系,在资金、资源、服务、传播等方面共同推动公共数字文化建设,形成共建共享的良好局面。

三是,聚焦发展问题与矛盾,缩小不同区域与群体间发展差距。发展过程中凸显的问题与矛盾,往往促使相关部门重新审视并调整公共数字文化政策;同时,这些问题的解决也为政策完善与行业进步提供了关键契机。因此,必须高度重视并着力解决发展中的突出问题,以促进自身持续发展。针对当前区域与群体间的发展差距,一方面,要加强政策引导,通过设立文化发展专项基金等支持措施,鼓励文化资源向欠发达地区以及老年人、残障人士、少数民族等群体倾斜。另一方面,要依托数字化技术,通过文化数字化传播焕发少数民族文化、乡土文化的活力与创造力,拓展其发展空间和市场边界,助力实现数字文化驱动的共同富裕。

本研究的不足之处在于:在LDA主题建模过程中采用后离散时间方法进行主题识别,虽避免了主题不对齐问题,但未能很好解决各时间窗内政策文本数量不等造成的主题识别准确度不优的问题,下一步

研究可以考虑采用更为先进的主题模型进行主题识别,完善相关研究过程。

参考文献:

- [1] 李放,饶丹,沈苏燕.“工具—主题—属性”框架下家庭养老支持政策量化评价研究[J].人口与发展,2023,29(3):84-95.
- [2] 王旭,李雨晴.我国个人信息保护政策文本量化研究及启示——基于“工具-效力-主题”三维框架的探析[J].情报科学,2023,41(1).
- [3] 吴鹏,张蒙蒙,索金琳.基于时空大数据演化的新冠肺炎疫情政策主题挖掘及层级扩散特征分析[J].情报理论与实践,2023,46(5):185-192+153.
- [4] MA C W, CHEN Q R. Spatial-temporal evolution pattern and optimization path of family education policy: An LDA thematic model approach[J]. Heliyon, 2023(7): e17460.
- [5] 胡吉明,钱玮,李雨薇,等.基于 LDA2Vec 的政策文本主题挖掘与结构化解析框架研究[J].情报科学,2021,39(10):11-17.
- [6] WRIGHT D, GRAY C. Culture is digital; and the shifting terrain of UK cultural policy[J]. International journal of cultural policy, 2022(7): 799-812.
- [7] 吴丹,张书田.融合创新视角下我国公共数字文化政策回溯、解读与展望[J].图书与情报,2021(1):45-52.
- [8] 陈庚,林嘉文.我国数字文化产业政策的演进脉络、阶段特征与发展趋势[J].深圳大学学报(人文社会科学版),2022,39(6):40-51.
- [9] 朱益平,金悦,樊丽珍.我国公共数字文化服务政策主题变迁与文本量化研究[J].图书馆建设,2024(1):74-86.
- [10] 马海群,崔文波,张涛.我国数据安全政策文本主题挖掘及其演化分析[J].现代情报,2024,44(8):28-38.
- [11] 苏婉,于森,程传阳.我国数据要素政策主题演化与识别分析[J].图书情报工作,2024,68(16):90-103.
- [12] 郭强,刘冬梅.基于 LDA 模型的农业农村科技创新政策扩散特征研究[J].中国软科学,2023(1):32-39.
- [13] WU Z Z, XIE P Y, ZHANG J M, et al. Tracing the trends of general construction and demolition waste research using LDA modeling combined with topic intensity[J]. Frontiers in public health, 2022, 10: 121-129.
- [14] 刘自强,王效岳,白如江.多维主题演化分析模型构建与实证研究[J].情报理论与实践,2017,40(3).
- [15] 霍朝光,董克,司湘云.国内外 LIS 学科主题热度演化分析与预测[J].图书情报知识,2021(2).
- [16] 王晰巍,李聪聪,孙哲,等.重大突发事件下网络舆情传播中社会群体主题演化图谱构建研究[J].图书与情报,2024(2):80-91.
- [17] WANG X F, ZHANG S, LIU Y Q, et al. How pharmaceutical innovation evolves: The path from science to technological development to marketable drugs[J]. Technological forecasting and social change, 2021, 167: 120698.
- [18] 王伟,高宁,徐玉婷,等.基于 LDA 的众筹项目在线评论主题动态演化分析[J].数据分析与知识发现,2021,5(10):103-123.
- [19] 周爱霞,严亚兰,查先进.基于神经网络词嵌入的大数据关注热点和词嵌入概貌比较研究[J].现代情报,2024,44(1).
- [20] 胡译文,王梦雅,韩雅蓉.基于 LDA2Vec-BERT 的新兴技术主题多维指标识别与演化分析研究——以颠覆性技术领域:区块链为例[J].现代情报,2024,44(9):42-58.
- [21] BLEI M D, NG Y A, JORDAN M I. Latent dirichlet allocation[J]. Journal of machine learning research, 2003, 3: 993-1022.
- [22] 高楠,高嘉骥,陈洪璞.新兴技术识别与演化路径分析方法研究——以集成电路领域为例[J].情报科学,2023,41(3):127-135+172.
- [23] 杨长进,罗仁杰,黄俊,等.产业政策对新能源汽车企业技术创新的影响——基于政策文本的 LDA 模型分析[J].科技进步与对策,2025,42(14):114-126.
- [24] WANG B, LIU S B, DING K, et al. Identifying technological topics and institution-topic distribution probability for patent competitive intelligence analysis: A case study in LTE technology[J]. Scientometrics, 2014(1): 685-704.
- [25] 曹玲静,张志强.政策信息学视角下政策文本量化方法研究进展[J].图书与情报,2022(6):70-82.
- [26] 张鑫,文奕,许海云,等. Prophet 预测-修正的主题强度演化模型——以干细胞领域为实证[J].图书情报工作,2020,64(8):78-92.

- [27] 曾江峰,黄泳潼,陈静怡,等.面向融合出版技术的主题演化研究[J].数字图书馆论坛,2023,19(4):9-18.
- [28] 曹玉梅,邢占军,王晓武.我国青年网络文化政策发展趋势研究——基于文献量化研究的视角[J].中国青年研究,2018(11):62-68.
- [29] BATES J M,GRANGER C W J. The combination of forecasts[J]. Journal of the operational research society,1969,20:451-468.
- [30] 阿布都热合曼·阿布都艾尼.基于博弈论组合赋权模型的跨境电商物流业务评价[J].统计与决策,2022,38(19):184-188.
- [31] 孙卫忠,王志波,高迎平,等.运用博弈论组合赋权法的图书馆藏优化研究——以河北工业大学图书馆为例[J].图书情报工作,2018,62(10):40-46.
- [32] 朱晓卉,胡彦蓉,刘洪久.基于微博核心实体的情感分析方法及引导机制研究[J].情报科学,2022,40(3):136-143+165.
- [33] 谢紫悦,陈雅,王丹.全景与远景:我国公共文化服务数字化建设进向分析[J].数字图书馆论坛,2023,19(5):17-25.
- [34] 王子健,完颜邓邓.政策工具视角下我国公共数字文化服务政策文本分析[J].图书馆建设,2023(2):117-130.
- [35] 彭丽徽,顾般若,洪闯.我国公共数字文化服务政策的脉络演进及文本量化分析[J].图书馆杂志,2024,43(6):27-39.
- [36] 温兴琦,陈思颀.我国数字文化产业政策工具的演变脉络与优化建议[J].三峡大学学报(人文社会科学版),2024,46(3):86-91.
- [37] 刘慧君,吴鹏.健康老龄化服务效能指标体系构建与实证测度[J].中国人口科学,2023(1):114-128.

Evolution of Topics on China's Public Digital Cultural Policies from the Perspective of "Strength-Content"

HOU Yanhui, ZHAO Wenjing, WANG Jiakun

(College of Economics and Management, Shandong University of Science and Technology, Qingdao, Shandong 266590, China)

Abstract: In order to analyze the topic distribution and evolution path of public digital cultural policies in China, effectively reveal the trends in policy changes, grasp the underlying mechanism of policy topic evolution, and provide theoretical support for policy formulation and implementation, this paper collects and examines the texts of public digital cultural policies in China, and uses LDA model to identify the topics and keywords within these policy texts. Additionally, a multi-feature integrated model representing topic strength is constructed to analyze the evolution of topic strength. Furthermore, the Word2Vec model is employed to obtain the most similar word set of topic keywords and analyze the evolution of topic content. The research finds that public digital cultural policies in China mainly encompass four topics: three major projects for the benefits of the people, equitable allocation, construction and security management, and the development of digital cultural industry. The evolution of these topics presents the characteristics such as periodic changes, science and technology driven trends, and a focus on development issues. Based on the results of topic evolution analysis, corresponding suggestions for the future construction of public digital culture are proposed.

Key words: public digital culture; topic strength; multi-feature fusion; topic content; topic evolution

(责任编辑:魏 霄)