

基于LDA模型的我国体育产业高质量发展的政策分析

俞宗宝

武汉体育学院 经济与管理学院

摘要: 自国务院2019年发布《国务院办公厅关于促进全民健身和体育消费推动体育产业高质量发展的意见》, 各省市陆续发布相关意见, 从而引发对于体育产业高质量发展政策现状的研究。本文结合LDA主题分析模型与主题社会网络等方法, 探讨了体育产业高质量发展政策文本的关注重点以及结构特征, 并得出结论。

关键词: 体育产业高质量发展政策; 文本挖掘; LDA模型; 主题社会网络

【DOI】10.12252/j.issn.2096-6288.2022.04.211

一、研究方法数据来源

(一) 研究方法

1. 文本挖掘

文本挖掘是指通过对大量非结构化文本集进行处理分析并得出其中隐含的信息、知识等涵盖计算机、数学与语言等学科领域知识的量化分析方法。由于政策文本内容繁杂, 因此对于政策文件的分析方法来说, 其分析方法具有独特性。本文在政策文本挖掘中所采用的方法与程序为: 首先, 将收集的政策文件划分为国家、省、地市州(后文中统一称为市)三个层级; 其次, 使用Python语言中自带的jieba库, 对各层级的政策文本进行停用词与保留词处理, 随后进行分词, 并对同义词进行合并, 得出各层级政策文本的总体词频数, 并绘制相应词云图; 最后, 根据关键词词典, 对政策文本进行关键词的词频统计。

2. LDA主题模型

LDA是一种能够将若干文档自动编码分类为一定数量的主题, 并将主题按照概率分布的形式给出的主题模型。于2003年, 由Blei首先提出LDA的相关概念, 并于后期提出了困惑度的概念, 将其作为评价主题分类好坏以及给出主题划分建议的标准⁷。

本文主要运用LDA模型, 挖掘体育产业高质量发展政策文本的主题特征, 以此确定各层级政策文本的关注重点。

α 与 β 分别控制一个迪利克雷分布, α 随机生成本文对应主题的多项式分布 θ , θ 随机生成主题 z , β 随即生成主题对应的词语的多项式分布 ϕ , 最后综合主题 z 与词语的多项式分布 ϕ , 生成词语 w , 最终生成 N 篇文档下的 K 个主题的 M 个词语。其中取主题数 $k=5$, 参数 α 和 β 值分别 0.1 和 0.01^5 , 因本文的政策本文总计 61 份, 而LDA主题模型只适用于大样本分析, 若分层进行LDA主题模型分析, 具有很大的偏差, 故此, 将各级政策文本统一分析, 确定体育产业高质量发展的总体主题。

3. 社会网络分析

社会网络分析(SNA)方法, 亦称之为结构分析法, 主要用于分析社会网络的关系结构及其属性。现今, 在图书情报领域, 已有很多研究将LDA主题模型与社会网络分析相结合, 例如酒店用户评价研究⁶, 政策文本计量研究⁸等。因此, 本文也采用同样的方式, 在LDA主题模型的基础上, 引入SNA构建主题与主题, 主题与特征词之间的社会网络, 从而探究主题与主题之间的关联关系, 单一主题下特征词的共现关系以及对LDA模型的主题聚类效果进行评价。

二、研究结果

(一) 政策文本高频词汇分析

本文在停用词和保留词处理的基础上, 对各层面政策文本进行了分词处理, 得出每个层面政策文本的总体词频数。而为了直观的展示不同层级政策的热点话题以及层级间政策的联系关系, 本文对高频词汇进行了汇总与排序, 并选取词频数最高的前30个词汇。

从三个层级中“支持”“发展”“建设”“鼓励”的词频均很高。表明各层级政府关于促进全民健身与体育消费推动体育产业高质量的政策的目标具有一致性, 而在三个层级中, 关键词词频排名的变化则反映出各个层级的侧重点有所不同。在国家层面, 主要提及了“户外运动”“冰雪”等, 从宏观上确定了体育产业高质量发展的大体方向; 省级层面, “冰雪”仍作为高频词汇, 表明冰雪产业的高质量发展是各省贯彻落实国家出台的体育产业高质量发展政策的一个集中体现, 此外, 与国家层面的高频词汇相比, 省级层面更加强调了“体育产业”一词, 表明省级层面的政策文本更加聚焦于体育产业, 突出体育产业的核心地位; 市级层面除与省级层面一致的高频词外, “体育部门”一词的词频较高, 表明体育部门是市级层面政策措施的执行主体。此外, “体育赛事活动”一词在市级层面的词频排名高于在国家层面与省级层面的词频排名, 表明各市着力打造体育

赛事活动,从而促进城市发展⁹。

“体育部门”一词的词频排名,从国家到市级层面逐级提升,表明在政策落地实施的过程中,对体育部门的职责进行了更为清晰明确的界定。例如,国办发(2019)43号文件在体教融合方面,提出支持学校与体育部门建立运动员共同培养机制。以游泳、田径等项目为试点,将教育部门主办的符合要求的赛事纳入运动员技术等级评定体系。浙政办发(2020)17号文件在此基础上,将项目扩展到游泳、田径、足球、篮球、排球、射击等,并提出了要开展青少年体育研学活动。绍政办发(2020)33号文件则提出支持社会力量参与开发并将体育基地、运动营地等纳入青少年研学基地,认定若干体育类中小学教育研学实践基地。随着社会力量的参与和体育部门职责的不断扩充,非政府性的体育组织应运而生¹⁰。因此,“体育组织”一词出现在了市级层面的前30个词汇之中。此外,从国家层面对于消费的提及到省市层面体育消费的关注度不断增加,以及“产业”一词的词频排名随层级下降而下降的同时“体育产业”一词的词频排名不断提高,体现了政策文本由宏观的战略与方向转变至微观的方式与方法的特征。

(二) 政策文本主题模型分析

1. 主题强度分析

主题强度是指聚类后的每个主题在语料库中占的相对分量,计算结果能够反映体育产业高质量发展政策对某些领域的关注程度,主题强度高代表该领域为政策的关注重点⁵。根据主题强度计算公式对三个层级政策的主题强度进行计算,并运用LDAvis可视化分析软件,直观呈现5个主题的分布情况

从总体来说,关于促进全民健身与体育消费推动体育产业高质量发展的政策文本能做到全面性与广泛性,但是仍有重点领域的突出强调。

2. 主题结构与困惑度分析

通过LDA,不仅能够确定主题词,还可以得到每个主题下的词汇分布结构。这种结构通过每个词汇的概率大小来表示概率越大,证明该词与主题之间的关系越紧密。在此基础上,通过主题困惑度计算公式(3),寻找困惑度图中极值点,最终科学地确定主题的数量。

1) 协同治理

“体育部门”“文旅部门”“住建部门”“广电局”是主题1中相关度最高的词频。所谓协同治理,是指涵盖多个层级,包括政府、市场、社会等,以实现公共目的的公共政策制定与管理过程与结构¹¹。因此,将主题一一定义为协同治理。例如,国家体育总局、

文化和旅游部、国家发展改革委等3部门联合印发《冰雪旅游发展行动计划(2021—2023年)》;云政办发(2020)22号文件在强化保障措施中提出,加强市场监管,应由省市场监管局牵头,省发展改革委、省体育局配合,均体现出协同治理的发展趋势日益强化。

2) 财政补助

将主题2命名为财政补助,“专项资金”、“资助”、“财政”、“经费”为其主要特征词,表明政策以专项资金以及经费等形式来对体育产业进行相应补助,从而促进体育产业的高质量发展。例如,鲁政办发(2021)18号文件中提出通过省级体育发展资金从而支持体育产业发展,鼓励各级设立体育产业发展专项资金。

3) 空间布局

将主题3命名为空间布局。其中,“街道”“格局”“地方”“集聚区”“落户”“落地”“空间布局”“市级”与主题3的相关度高,表明政策重视体育在活化城市空间、形成产业的空间集聚以及传播城市影响力等方面的作用。例如,在《户外运动产业发展规划(2022—2025年)》中提出,要以京津冀为核心区域,以2022北京冬奥会的成功举办为契机,加快建设京张体育文化旅游带,推动东北、华北、西北冰雪运动发展,打造具有国际影响力的北方冰雪运动引领区。

4) 跨界融合

“滑雪场”“度假区”“智能”“高端”为其主要特征词,由此可体现出政策文本重视体育产业的跨界融合,特别是体育产业尤其是冰雪产业与旅游、科技服务业融合发展。国办发(2019)43号中提出,实施“体育+”行动,促进体育与健康、文化、旅游、养老、科技等融合发展,也可体现跨界融合已成为体育产业高质量发展的一个重要“新常态”¹²。

(三) 主题社会网络分析结果

本部分在基于LDA主题分析模型的基础上,选取每个主题高频词汇中最能体现该主题特征的5-8个特征词,总计24个特征词,并根据主题强度,分配每个主题下特征词的数量,结果如表3所示,并依据这些特征词构建社会网络。

具体来说,本部分使用Python语言构建关键词共现矩阵,并通过共现网络考察主题与主题之间的关联关系,单一主题下特征词的共现关系以及对于LDA模型的主题聚类效果进行评价。

主题间连线线条的粗细度与对应节点主题共同出现的数量呈正相关。在该主题社会网络图中可以看出,财

政补助与跨界融合之间的线条粗细差异程度小,即体育产业高质量发展政策对财政补助与跨界融合这两个主题没有明显的偏好;而空间布局与协同治理这两个主题由很粗的线条连接,即空间布局与协同治理这两个主题之间的联系尤为密切,体育产业高质量发展政策文本在考虑空间布局时,通常会涉及协同治理,其中,体育部门与街道之间的连线最粗,即关系最为密切,这集中体现了体育产业高质量发展的政策一定要落于实地,而不是束之高阁。

三、结论与启示

(一) 结论

本文通过对高频词汇的词频统计得出,国家、省、市三个层面的政策文本在总体目标上具有一致性,三个层面的政策从制定到落实是一以贯之的。但是不同层级的政策文本的关注重点各不相同,国家层面的政策主要提及了“户外运动”“冰雪”等,从宏观上确定了体育产业高质量发展的大体方向;在省级层面,更加关注结合政策实施的方针与策略,特别是在冰雪产业的高质量发展方面,各省做出了结合地区特色的具体规划,此外,体育赛事活动的提及说明了省级层面的政策亦有自己的关注重点;而在市级层面,对于各类省市级赛事活动的突出强调,更多体现的是结合当地特征的微观方式方法。

(二) 启示

针对现有体育产业财政补助的问题,亟待政府部门进行改善。改善措施包括:1.顶层设计,构建基于体育产业高质量发展的税收优惠政策框架体系,从总体上确定体育产业高质量发展的税收优惠政策的实施原则和重点支持对象,并且通过部门间的协同治理,将政策一以贯之,从而解决税收优惠政策法律层级较低,缺乏稳定性与持续性的问题;2.加大体育产业的资金投入规模,并实现多种补贴方式并举,政府除了直接的“点对点”对体育企业,特别是融入科技创新的体育企业进行补贴外,还可以与专业的投资企业进行合作,由投资企业来对补贴对象进行系统性地评估,亦可由政府打造线上融资平台,简化中间程序,提升资本的运营效率;3.由于各省市经济发展状况差距较大,税收优惠政策须因时因地因企业而异,可根据企业生命周期理论来对各地的体育市场现状进行归类,从而选择相对应的政策。例如在初创期与衰退期,由于企业的投资风险较大,基本持观望态度,政府须加大财政投入力度,激励企业投资体育产业¹⁵。

针对体育产业与科技服务业融合发展方面存在的问

题,可以从这些方面进行改进完善:1.加强顶层设计,国家应出台科技助力体育产业特别是数字体育方面的指导文件,制定科技助力体育产业翔实的发展规划;2.基础设施较差的地区除了要根据当地实际情况,来加大对基础设施的建设外,一定要利用政策文本中对于空间布局与协同治理重点关注的趋势以及自身的独特优势,例如生态环境、市场需求量大等,从而与设施建设较好的地区展开深度合作,完善基础设施;3.加强人才培养,数字体育人才的培养需要理论与实践相结合,既有较高的理论知识水平又具备较强的实际操作能力,既能掌握数字科技的关键技术又对体育各领域有深入了解¹³,因此,各大高校应该开设相关领域的课程,使学生掌握数字化技术的理论知识,并且企业应积极与高校开展产学研合作交流,从而使学生能够掌握较强的实际操作技能;4.财政政策应重点支持数字体育,智能体育企业,科技与体育融合的新兴产业。此外,各地政府应多开展展览会,高端赛事,增加科技产品与智能体育企业的曝光度。5.二十大中提及,广泛开展全民健身活动,加强青少年体育工作,促进群众体育和竞技体育全面发展。因此,科技助力体育产业特别是数字体育应重点关注全民健身、健康中国、竞技体育和体教融合等方向。

参考文献

- [1]张凯,陶玉流.我国体育产业政策的演变趋势与特征——基于2012—2019年的政策文本分析[J].体育科研,2021,42(05):83-89.
- [2]徐开娟,黄海燕,廉涛,李刚,任波.我国体育产业高质量发展的路径与关键问题[J].上海体育学院学报,2019,43(04):29-37.
- [3]芦胜男,刘冬磊,王子朴.基于政策工具视角下我国体育消费政策分析——基于37份国家政策文本的内容分析(2014—2019)[J].武汉体育学院学报,2021,55(01):51-58.
- [4]田园,宫婷婷.基于LDA模型的在线教学需求数据主题挖掘研究[J].情报科学,2021,39(09):110-116.
- [5]李磊,李梓阁.基于LDA主题模型的自贸区治理政策文本聚类分析——以辽宁自贸区为例[J].吉首大学学报(社会科学版),2021,42(02):23-34.
- [6]池毛毛,潘美钰,王伟军.共享住宿与酒店用户评论文本的跨平台比较研究:基于LDA的主题社会网络和情感分析[J].图书情报工作,2021,65(02):107-116.