

◁ 信息管理 ▷

健康医疗大数据应用发展评价模型构建研究

韩丽颖¹, 迟蔚蔚¹, 王玥¹, 周晗²

(1. 山东大学公共卫生学院, 济南市 250100; 2. 山东省疾病预防控制中心, 济南市 250100)

【摘要】 目的 构建山东省医疗机构健康医疗大数据应用发展评价模型。方法 通过文献研究方法, 基于DPSIR理论模型, 运用专家咨询法对指标的重要性进行打分, 问卷调查法、熵权TOPSIS计算完成权重估计, 构建健康医疗大数据应用发展评价体系。以山东省为例, 对该评价模型进行实证研究。结果 构建了6个维度、25个二级指标、57个三级指标的山东省健康医疗大数据应用发展评价模型, 并对省域内34家医院进行评分。结论 健康医疗大数据应用发展评价指标模型具有科学性、实践性及可推广性, 能较为科学合理地评价山东省医疗机构健康医疗大数据应用发展情况。

【关键词】 医疗机构; 健康医疗大数据; 应用发展评价模型; 熵权TOPSIS**【中图分类号】** R197 **【文献标识码】** B **【文章编号】** 1672-4232(2024)06-0103-04**【DOI编码】** 10.3969/j.issn.1672-4232.2024.06.027

发展大数据产业是国家的重要战略目标^[1], 医疗大数据作为大数据的重要组成部分^[2], 正越来越受到国家和社会的关注。发达国家也十分重视对医院信息化以及健康医疗大数据应用发展水平的评价工作^[3]。

目前国内外对于医疗机构的信息化建设及健康医疗大数据应用发展的研究争相开展^[4], 且主要从资金投入、人才队伍规模、基础设施配备、信息技术的应用水平以及信息化绩效等维度展开评价^[5-6], 但缺乏对数据资源质量、数据资源的共享与利用方面的考察, 且缺乏成熟的评价模型对评价体系进行优化整合^[7-8]。因此, 迫切需要建立一套综合性评价指标模型, 能够全面反映健康医疗大数据的发展现状, 将评价指标与成熟评价模型相结合, 具有广泛的应用价值, 以推动医疗卫生机构大数据应用的发展。

1 研究方法

1.1 文献研究法

通过以关键词“大数据发展”、“评价模型”等综合查阅相关文献, 深入了解健康医疗大数据的内涵、应用发展、指标体系构建以及综合评价方法, 为本研究提供理论支持和方法借鉴。

1.2 专家咨询法

由来自山东省10个地市的18位卫生及大数据领域专家和卫生行政部门工作人员组成专家组。采用现场座谈和线上会议。基于专家评分和意见, 对指标体系进行增、删、改。通过专家积极程度和权威程度验证专家组的可靠性^[9-10]。

1.3 问卷调查法

设计《医疗机构健康医疗大数据应用发展测评问卷》, 向40家医疗机构发放问卷40份, 回收有效问卷

34份。依据问卷内容对其医疗大数据应用发展情况进行评价。

1.4 熵权法

评价指标权重采用熵权法, 通过三级指标体系逐层加权。首先, 对健康医疗大数据应用发展评价的特征指标进行归一化处理, 然后运用信息熵概念量化每个指标的不确定性。采用逐层加权方法, 考虑不同级别指标对总评价的贡献, 确保多层次综合评估。

2 评价指标模型构建

2.1 构建思路

本研究探索性地将DPSIR模型引入到健康医疗大数据应用发展评价指标体系的构建中, 思路为选定理论模型、形成指标池、指标筛选、构建体系框架、计算指标权重、模型实证研究^[11]。首先依据驱动力、压力、状态、影响及响应五个准则层构建指标体系; 在对国内外相关研究文献归纳总结的基础上, 结合我国健康医疗大数据发展的现状和健康医疗大数据政策文件^[12-13], 梳理健康医疗大数据应用发展的各维度指标, 形成指标池。通过访谈、调研、问卷等方式获取山东省医疗卫生机构的数据, 分析挖掘并确定具体指标, 构建评价体系框架。采用熵权TOPSIS计算权重, 最终建立健康医疗大数据应用发展评价体系。以山东省医疗机构为案例进行实证研究, 验证体系科学性和有效性, 探索在实践中的应用效果和价值。

2.2 指标体系

健康医疗大数据应用发展评价体系包括资金投入、人力资源配置、基础设施、软件技术、数据资源和政策6个维度, 医院收入、信息化人才队伍规模、云计算服务投入、医院规模、机房配置等25个二级指标, 近3

年信息化建设总投入经费等57个三级指标,综合评估了山东省医疗机构健康医疗大数据应用发展水平及其主要因素,详细指标体系见图1。

2.3 指标权重

本研究采用熵权法完成权重的估计^[14-15]。该方法的步骤开始于数据预处理,包括选取合适的样本和指标,以及处理数据缺失和异常。数据标准化使用min-max极差法,将数据转换到统一尺度,消除量纲影响。

接着,通过计算比重: $y_{ij} = \frac{x'_{ij}}{\sum_{i=1}^m x'_{ij}}$ ($0 \leq y_{ij} \leq 1$);来确定每个样本在各指标下的分布。基于比重,进一步计算每个指标的信息熵: $e_j = -\frac{(\sum_{i=1}^m p_{ij} \times \ln p_{ij})}{\ln m}$,以及冗余度: $d_j = 1 - e_j$;再由冗余度计算指标的权重: $w_j = \frac{d_j}{\sum_{j=1}^m d_j}$,从而计算出每个样本的综合评价得分: $S = \sum_{j=1}^m w_j \times y_{ij}$ 。

3 结果

3.1 数据预处理结果

剔除缺失值较多的指标,并对无序多分类及有序多分类变量进行转换,引入哑变量将不能够定量处理的变量量化,并根据分类指标定量转换进行分级评价。

3.2 评价指标权重

经指标筛选及数据预处理,形成山东省医疗机构健康医疗大数据应用发展评价三级指标体系。经熵权法逐层加权获得三级指标权重,权重之间呈现一定的差异。结果见表1。

3.3 研究对象得分

由加权求和计算得出山东省参与调研的多家医疗机构健康医疗大数据应用发展评价得分,可以评价得分最高的医院得分为基准,进而分别计算其他医疗机构的相对比值,计算得分。

4 讨论

为解决当前医疗大数据应用发展评价的挑战,本研究基于实地调研和问卷等手段,由山东省医疗大数据管理中心收集数据,构建了一个科学且适用的评级指标体系。该体系结合了《国务院办公厅关于促进和规范健康医疗大数据应用发展的指导意见》和《“健康中国2030”规划纲要》,确保了指标的实用性与科学性,同时从因果视角整理,使指标设置具有学术依据。

通过专家论证,对指标体系进行了完善并设置了评分标准,确保指标间结构合理。利用数据驱动的权重计算,构建了目标导向的指标权重,与“互联网+医疗健康”发展目标相结合,使用熵权TOPSIS法评估实际与理想情况的差异,识别阻碍目标实现的短板,以促进健康医疗大数据应用发展。

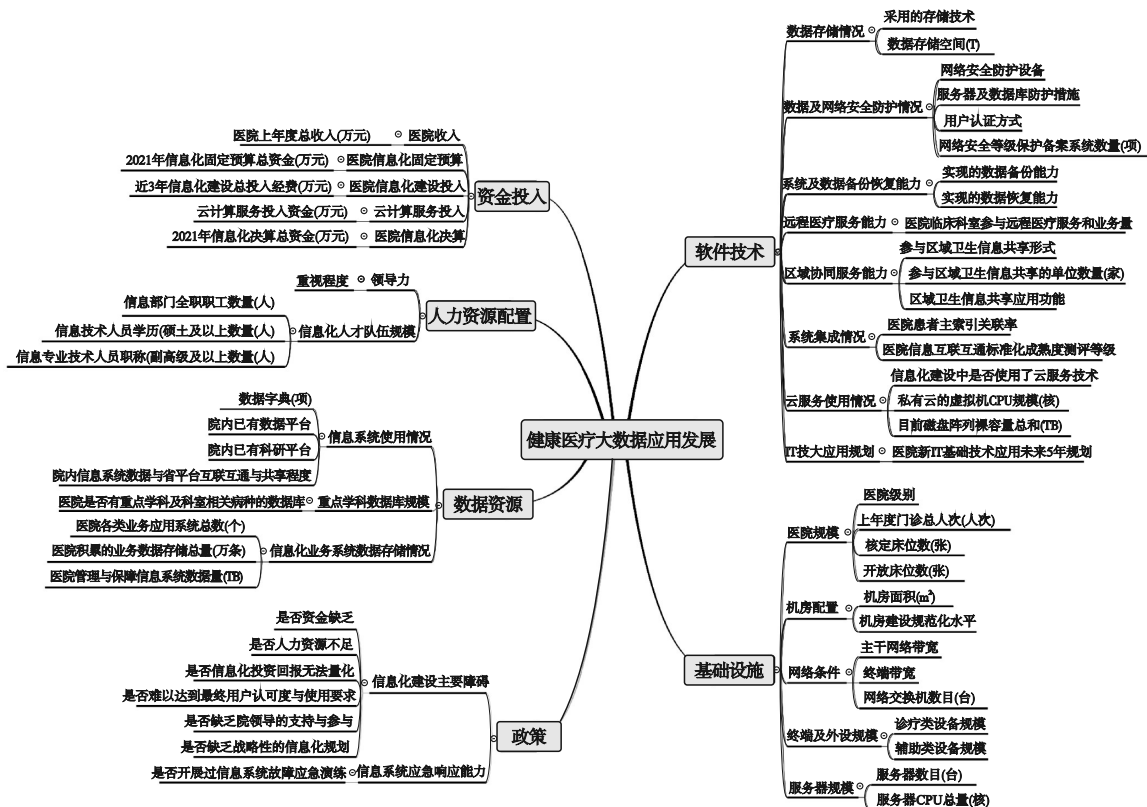


图1 健康医疗大数据应用发展三级指标体系

表1 山东省医疗机构健康医疗大数据应用发展评价指标权重

一级指标(权重)	二级指标(权重)	三级指标(权重)
资金投入(0.067)	医院收入(0.015)	医院上年度总收入(0.015)
	医院信息化固定预算(0.010)	2021年信息化固定预算总资金(0.010)
	医院信息化建设投入(0.001)	近3年信息化建设总投入经费(0.001)
	云计算服务投入(0.037)	云计算服务投入资金(0.037)
人力资源配置(0.034)	医院信息化决算(0.004)	2021年信息化决算总资金(0.004)
	信息化人才队伍规模(0.030)	信息部门全职职工数量(0.007)
		信息技术人员硕士及以上数量(0.011)
基础设施(0.146)	领导力(0.004)	信息技术人员副高级及以上数量(0.012)
	医院规模(0.032)	重视程度(0.004)
		医院级别(0.003)
		上年度门诊总人次(0.012)
		核定床位数(0.009)
		开放床位数(0.008)
	机房配置(0.020)	机房面积(0.011)
		机房建设规范化水平(0.009)
	网络条件(0.010)	主干网络带宽(0.002)
		终端带宽(0.004)
终端及外设规模(0.013)		网络交换机数目(0.004)
		诊断类设备规模(0.006)
		辅助类设备规模(0.007)
服务器规模(0.045)		服务器数目(0.013)
		服务器CPU总量(0.032)
数据存储情况(0.027)		采用的存储技术(0.012)
		数据存储空间(0.015)
软件技术(0.308)	数据及网络安全防护情况(0.045)	网络安全防护设备(0.001)
		服务器及数据库防护措施(0.027)
		用户认证方式(0.010)
		网络安全等级保护备案系统数量(0.007)
	系统及数据备份恢复能力(0.005)	实现的数据备份能力(0.002)
		实现的数据恢复能力(0.003)
	远程医疗服务能力(0.003)	医院临床科室参与远程医疗服务和业务量(0.003)
	区域协同服务能力(0.072)	参与区域卫生信息共享形式(0.005)
		参与区域卫生信息共享的单位数(0.032)
		区域卫生信息共享应用功能(0.035)
系统集成情况(0.040)	医院患者主索引关联(0.023)	
云服务使用情况(0.139)		医院信息互联互通标准化成熟度测评等级(0.017)
		信息化建设中是否使用了云服务技术(0.032)
		私有云的虚拟机CPU规模(0.037)
		目前磁盘阵列裸容量总和(0.006)
		是否有云计算部署(0.032)
数据资源(0.261)	IT技术应用规划(0.004)	使用云计算前是否对厂商提出认证需求(0.032)
	信息系统使用情况(0.125)	医院新IT基础技术应用未来5年规划(0.004)
		数据字典(0.074)
		院内已有数据平台数量(0.013)
		院内已有科研平台数量(0.027)
		院内信息系统数据与省平台互联互通与共享(0.011)
	重点学科数据库规模(0.048)	医院是否有重点学科及科室相关病种的数据库(0.048)
	业务系统数据存储情况(0.088)	医院各类业务应用系统总数(0.013)
		医院积累的业务数据存储总量(0.034)
		医院管理与保障信息系统数据量(0.041)
政策(0.184)	信息系统应急响应能力(0.002)	是否开展过信息系统故障应急演练(0.002)
	信息化建设主要障碍(0.004)	是否资金缺乏(0.044)
		是否人力资源不足(0.069)
		是否信息化投资回报无法量化(0.035)
		是否难以达到最终用户认可度与使用要求(0.019)
		是否缺乏院领导的支持与参与(0.001)
	是否缺乏战略性的信息化规划(0.013)	

本研究的创新性在于首次将DPSIR模型应用于医疗大数据应用发展评价,采用熵权法进行权重计算,弥补了主观赋权法的不足,提高了权重分配的客观性

和真实性。这为全国各级医疗机构及卫健委提供了有益参考。

尽管在实证研究中存在一定局限性,(下转第132页)