

以诊疗数据挖掘为指导的三甲医院发展策略研究^{*}

高 瞻

高 艳

韩玖阳 赵 亮

(十堰市太和医院 十堰 442000) (十堰市人民医院 十堰 442000) (十堰市太和医院 十堰 442000)

〔摘要〕 基于医院患者住院数据构建科室综合能力分析模型, 阐述数据处理与分析方法、模型构建路径, 对结果进行横向、纵向分析, 提出科室发展策略并进行案例分析, 为医院管理者制定最优发展策略提供参考。

〔关键词〕 科室运营; 科室演化; 发展策略

〔中图分类号〕 R-058 〔文献标识码〕 A 〔DOI〕 10.3969/j.issn.1673-6036.2022.02.008

Study on the Development Strategy of Grade III Level A Hospitals Based on the Mining of Diagnosis and Treatment Data GAO Zhan, Shiyan Taihe Hospital, Shiyan 442000, China; GAO Yan, Shiyan Renmin Hospital, Shiyan 442000, China; HAN Jiuyang, ZHAO Liang, Shiyan Taihe Hospital, Shiyan 442000, China

〔Abstract〕 The comprehensive ability analysis model of department is built based on the data of patients in hospital. The paper expounds the data processing and analysis methods and the building path of the model, analyzes the results horizontally and vertically, puts forward the department development strategy and case analysis, and provides references for hospital managers to formulate the optimal development strategy.

〔Keywords〕 department operation; department evolution; development strategy

1 引言

1.1 研究背景

新医改日益深化, 私立和专科医院兴起, 公立

〔修回日期〕 2021-05-31

〔作者简介〕 高瞻, 硕士, 初级职称, 发表论文 7 篇; 通讯作者: 赵亮, 博士, 副研究员。

〔基金项目〕 湖北医药学院自由探索基金项目“室间隔缺损突变基因所致蛋白质功能改变分析及自愈关键指标挖掘”(项目编号: FDFR201805); 十堰市太和医院院级基金项目“肝脏影像智能诊断系统实现及三维可视化研究”(项目编号: 2020JJXM013)。

医院发展面临挑战。目前大型医院尤其是大型三甲医院科室众多, 向下划分的病区更加繁杂, 事务性工作主要由科主任管理, 高层管理者较难精准了解不同科室、病区特点及情况, 导致医院发展战略制定和实施在不同程度上受阻。近年来基于医院临床大数据, 通过机器学习方法挖掘数据潜在特征, 建立分析模型方面已取得显著成效。吴曼琪、潘习龙和程迪等^[1]利用临床大数据设计医疗质量监控辅助系统, 从环节上把握影响医疗质量与安全的关键点与关键事件, 全面提升医院医疗质量管理水平。赵娜、王吉善和王圣友等^[2]对医院评审评价结果关联规则进行分析, 发现医院管理中应重视建立各指标之间的关联性。在医院科室发展策略研究方面, 有学者^[3-4]运用数据包络分析 (Data Envelopment A-

nalysis, DEA) 方法对医院临床科室数据进行分析, 为医院合理规划临床科室配置及进一步提高资源利用率提供参考。陈伟和王忠^[5]应用加权综合指数法考核临床科室并提出针对性的发展建议, 为医院管理提供依据。

1.2 研究目的

传统医院科室(病区)人力、物资配备及发展方案制定主要是按收治患者情况被动调整, 医院管理者较难宏观把握综合发展的平衡点。以患者特征为基础, 结合医院总体发展战略布局, 将不同科室资源配备由被动变为主动, 制定切合实际的精细化发展策略是医院管理者亟待思考和解决的关键问题之一。本文以住院患者大数据为基础, 分析和挖掘各科室在不同维度上的特征信息, 基于上述特征信息利用聚类算法构建科室类群分析模型。根据类群特点将科室划分为发展型、全能型、服务型和技术型 4 类, 针对不同类型科室探究相应发展策略, 为医院科室管理和发展战略提供指导。

2 资料与方法

2.1 数据来源

经过多年发展, 目前各医院信息系统已积累大量临床、检验、影像以及患者数据等。其中患者数据主要包含患者基本信息(如年龄、性别、职业、地区、病史等)、住院信息(如住院时间、出院时间、住院科室、诊断疾病、用药信息、费用等)、手术信息(如手术时间、名称、级别等), 挖掘分析上述数据信息将为医院运营管理优化、患者疾病精准诊疗分析、科室个性化服务等提供依据。本文以某大型三甲医院现有医院信息系统(Hospital Information System, HIS), 检验信息系统(Laboratory Information System, LIS), 电子病历(Electronic Medical Records, EMR)等系统数据库为基础, 提取近几年住院患者数据共计 33 万余条, 见表 1。根据医院运营管理的最小粒度, 以科室(病区)为单位共计临床科室 79 个。

表 1 住院患者数据

信息类型	属性名称	属性说明
患者基本信息	Patient_ id	患者 ID
	P_ SEX	性别
	P_ BIRTHDAY	出生日期
	P_ Address	住址
	Insurance_ type	医保类型
	P_ occupation	职业
患者住院信息	Dept_ code	科室病区编码
	Dept_ name	科室病区名称
	ICD10_ code	主要疾病编码
	ICD10_ name	主要疾病名称
	IN_ date	住院时间
	out_ date	出院时间
	Charge_ Total	总费用
	Charge_ Insurance	医保报销费用
患者手术信息	Operation_ code	手术编码
	Operation_ name	手术名称
	Operation_ time	手术时间
	Operation_ level	手术级别

2.2 数据处理与分析

医疗行业具有一定特殊性, 公立医院发展不能过于追求营利性, 其基本医疗服务的公益性才是根本^[6]。因此以利益收入单因素衡量部门综合运营能力的模型并不适用于医院科室分析, 需要选取多项指标体现不同科室综合能力和特点。收入高低是科室发展的关键驱动力之一, 而接收患者数量多少及患者住院时长直接影响科室资源和人力配备情况。此外考虑到科室技术水平和综合业务能力, 接收患者所患疾病种类的广度和手术级别(分为 4 个等级, 4 级手术过程最复杂、技术难度最大、风险程度最高)的高低是重要因素。本文以住院患者数据为基础, 统计科室每年费用总收入、接收患者总量、患者住院总时长、收治患者疾病种类数、接收患者所做 4 级手术总量 5 个特征作为科室病区数据挖掘的价值指标, 构建科室综合能力分析模型。挖掘分析前需进行数据预处理, 包括删除关键字段缺失和异常值的数据; 患者住院总时长采用出院时间与住院时间差值; 同一患者住院期间多次手术, 只统计手术等级最高的 1 次。5 个指标数据取值范围差异较大, 为消除数量级带来的影响, 对数据进行

标准化处理，见公式（1）、表2。

$$x'_i = \frac{x_i - Mean(R_i)}{Std(R_i)}$$

(1)

其中 x_i 表示第 i 项特征值， $Mean(R_i)$ 为科室数据的第 i 项特征在其参考范围内的所有样本期望值， $Std(R_i)$ 为其在参考范围内的所有样本标准差。

表2 科室标准化后的数据示例

科室病区编码	住院患者总量	住院总时长	疾病种类数	4级手术总量	收入总费用
1020902	0.630354	0.968512	0.209458	4.986096	1.521743
1021103	-0.745779	-0.571249	-0.514689	2.079627	0.809687
1021202	-1.330707	-2.124928	-2.087835	1.121781	-1.501759
1021602	-0.237854	0.070188	1.433016	-0.480119	-0.234186
...	...	...	...	...	...
1022202	0.157043	-1.355588	0.500781	0.355697	-0.864963
1024002	-0.749312	-0.955800	-0.439777	0.120351	-0.392706

2.3 模型构建

2.3.1 聚类分群 科室综合能力分析模型构建主要分为两部分：根据科室5个指标数据进行聚类分群；分析每个类群特征，将科室划分成相应类型。科室类群划分采用机器学习算法K-Means，聚类数量K值可以结合对实际业务的理解和分析来确定，也可以选取多个K值，采用手肘法（Elbow Method）和轮廓系数（Silhouette Coefficient）^[7-8]方法找到聚类性能最优的聚类值K_{optimal}。其中手肘法是一种基于误差平方和（Sum of Squared Error, SSE）的K-均值聚类选择算法，该算法核心思想是随着聚类数

K的不断增大划分会更加详细，每个分组（簇）的聚合程度会不断地提高而SSE会逐渐减小，在达到某个临界点即“手肘”点时，畸变程度会得到极大改善，之后缓慢下降，这个临界点被认为是聚类性能较好的点^[9]。以十堰市太和医院近几年住院患者数据为基础，采用手肘法选取K值，计算结果显示数据聚类K值的表现基本相同。选取与当前医院情况最相近的2019年住院患者数据为例展开分析，在K=4时首次出现曲率最大的转折点即最优聚类值K_{optimal}=4。此时由聚类算法计算取得各类数量及类群中心点，见表3。

表3 医院各科室病区聚类结果

类别	科室数量（个）	聚类中心点				
		患者总量	住院总时长	疾病种类数	4级手术总量	收入总费用
类1	31	0.021	0.237	0.453	-0.279	0.145
类2	8	2.061	1.966	1.046	-0.174	0.349
类3	30	-0.673	-0.848	-0.837	-0.392	-0.829
类4	10	0.302	0.239	0.270	2.179	1.759

2.3.2 科室类群划分 针对2019年住院患者数据聚类结果进行特征分析并绘制雷达图，图中以不同符号标记不同类别，见图1。其中类别1科室数量在医院中占比最大，5项指标都处于中等偏上水平，既没有突出优势特征也没有明显弱势特征，在收治患者数量一定时，患者疾病种类最为复杂，是体现医院综合能力的关键，因此将其定义为全能型科室；类别2在患者总量、住院时长上具有

明显优势，手术难度和费用方面并不突出，如儿科、产科病区等，患者疾病种类相对单一、手术级别较低，但患者数量、陪护人员较多，人流量大，体现出该类科室具有一定服务与公益性质，定义为服务型科室；类别3在5项指标上均具有明显劣势，但科室数量在医院中占比较大，该类科室通常具有较强发展潜力，将其定义为发展型科室；类别4在4级手术量和收入总费用方面具有明显优势，

对医院绩效有突出贡献，但面临的患者疾病最为复杂，手术难度较大，为患者配备的相关医疗资源更全面，对核心医生技术水平要求较高，定义为技术型科室。

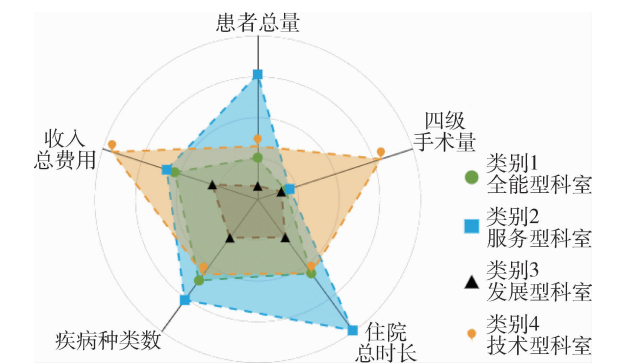


图 1 科室特征分析

3 结果与分析

3.1 横向分析

3.1.1 总体情况 通过科室综合能力分析模型，管理者可以精准掌握每个科室在医院中的类别定位，从宏观上了解不同类别科室特点，进而结合实际情况制定有针对性的发展方案。根据各科室类型特征，从重要性和发展潜力两方面对各类科室进行排序，见表 4。技术型和服务型科室在医院重要性上占据最关键地位，直接关系到医院绩效、技术影响力、收治患者情况等；全能型和发展型科室在医院中占比最大，其在各项指标上的不足反映出其具备较大发展潜力。根据不同类型科室特征提出 4 项发展策略。

表 4 医院各科室聚类结果排名

科室类型	数量占比（%）	重要性排序	发展潜力排序	典型科室病区
技术型	13	1	3	心血管内科病区、神经肿瘤疾病诊疗中心病区、泌尿男科等
服务型	10	2	4	儿科病区、产科、耳鼻咽喉科等
全能型	39	3	2	消化内科病区、呼吸与危重症医学科病区、感染科等
发展型	38	4	1	内分泌风湿病科、儿童康复中心、核医学科等

3.1.2 技术型科室 一方面，培养技术精深的专家，通过特殊患者治疗病例树立医生形象与医院品牌特色，不断扩大影响力；另一方面，注重青年骨干人才培养，学习国内外先进技术，推动科室持续发展与技术不断创新。

3.1.3 服务型科室 首先，应明确医院不能过于依靠该类科室提升绩效；其次，需要花费更多精力关注患者服务（如提供优质病房条件、患者切身关怀等），当医务人员和医疗资源有限的情况下可适当控制患者接收数量，从而保证服务品质和患者就医体验。最后，借助该类科室人流量大的优势宣传医院技术优势和相关惠民政策，建立医院良好口碑，从而吸引更多患者。

3.1.4 全能型科室 一是引入基础知识更广博和经验更丰富的医生，向全科医生方向发展以应对各类杂症，节省专科医生培养成本。二是通过

引进高水平人才、专家，增强医院对特大、特难疾病患者的收治能力，将部分科室发展成技术型科室。三是通过投入先进设备或建立专病专家组等手段吸引更多患者，将部分科室发展成服务型科室。

3.1.5 发展型科室 主要分为两种情况，一是科室收治的患者疾病罕见、数量稀少，但在医院战略布局上必不可少，保持在同等级医院的一般水平即可。二是科室发展潜力巨大，需要结合具体科室情况具体分析，投入相应人力、物力资源以供其发展，这类病区会随着医院的发展和战略布局进行改进和转变，表现为与全能型科室转化频繁。

3.2 纵向分析

采用上述模型对近几年医院科室数据进行分析并绘制科室演变图，见图 2。

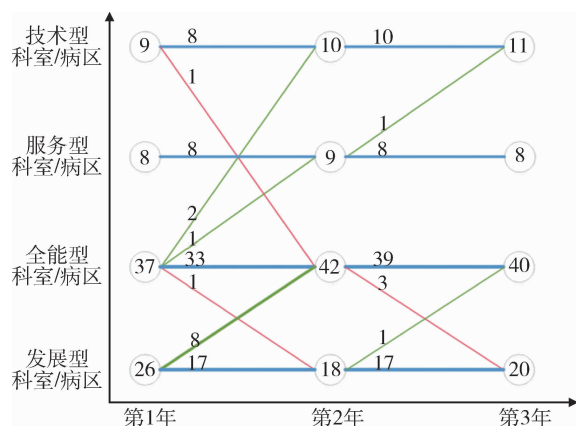


图 2 科室能力演变

注：结点上数字代表该类型科室数量，连线代表科室演化情况，线上数字代表转化数量。

以上述科室重要性排序为依据，从连线走势上看，连线斜率为正代表医院科室向好的方向积极发展，为负代表其发展在一定程度上受阻或衰退，平行演变则为稳定发展。从全局上看，全能型和发展型科室转化相对比较活跃，而技术型和服务型科室基本保持稳定态势。医院近几年有较多向上发展的科室，仅有 5 个科室出现下滑，大多数科室没有太大变化，总体上维持稳定态势。全能型和发展型科室在医院中占比达到 70% 以上，发展这两类科室能够有效提升医院综合实力。根据科室特征和能力演变分析，全能型向服务型科室转化的核心是发展患者数量，而全能型向技术型科室转化的核心是提升 4 级手术量，该现象提醒医院管理者可以在全能型科室中挑选最具潜力的科室，从上述两方面着手进行布局；发展型与全能型科室特征相似，均没有突出优势，但后者 5 项指标均明显优于前者，因此相互转化活跃，该现象提醒医院管理者可以分别挑选一一对应的发展型与全能型科室，共享全能型科室发展理念，为发展型科室树立学习标杆，相互帮扶、共赢。

4 结语

本文通过挖掘住院患者数据信息汇集成科室特征，采用聚类算法构建科室综合能力分析模型，进而将医院科室划分为发展型、全能型、服务型和技术型 4 种类型。针对不同类型科室制定相应发展策略，从横向和纵向两个角度分析医院发展情况。从宏

观上看服务型和技术型科室发展相对稳定，说明医院在口碑树立和品牌影响力上有持续投入，但其转化相对稳定，发展动力略显不足；全能型和发展型科室转化活跃，但绝大部分科室仍保持平行发展，其中除部分科室受到客观原因影响发展受限外，也存在一些科室由于没有找准自身定位、科室类别特征模糊，错失发展良机。通过对科室精准定位，深入了解各类科室特点，变被动调整为主动介入，使各类型科室保持持续快速的深化发展。本研究存在一定局限性：一方面，在特征挖掘时只选取收治患者数量、住院时长、疾病种类、住院花费、手术难度、手术数量作为衡量指标，其中手术难度也仅考虑 4 级手术，忽略了手术质量、治疗效果等关键临床指标；另一方面，科室发展策略分析与探究只是基于宏观数据，未结合具体情况深入展开，如科室医生及床位数量、各科室协作关系、区域性疾病情况等。在实际应用中管理者需在此基础上进行深入细致研究，制定符合本院实际和特色的最优发展策略。

参考文献

- 1 吴曼琪, 潘习龙, 程迪, 等. 基于临床大数据的医疗质量监控辅助系统探讨 [J]. 中华医院管理杂志, 2020, 36 (1): 53-55.
- 2 赵娜, 王吉善, 王圣友, 等. 基于关联规则的综合医院评审评价数据挖掘及管理启示 [J]. 中华医院管理杂志, 2020, 36 (8): 687-691.
- 3 胡鹏, 王志琳. 基于数据包络分析的医院临床科室运营效率研究 [J]. 中华医院管理杂志, 2019, 35 (1): 36-40.
- 4 杨毅, 周星宇, 陈英耀, 等. 基于数据包络分析的 H 省 MRI 科室资源配置效率研究 [J]. 中国医院管理, 2019, 39 (6): 22-24.
- 5 陈伟, 王忠. 综合评价方法在临床科室医疗质量评价中的应用 [J]. 上海交通大学学报 (医学版), 2013, 33 (7): 1018.
- 6 蔡进, 谭剑, 康静. 基于波士顿矩阵的公立医院科室业务发展战略研究 [J]. 医学与社会, 2019, 32 (5): 51-55.
- 7 Syakur M A, Khotimah B K, Rochman E M S, et al. Integration K - Means Clustering Method and Elbow Method For Identification of The Best Customer Profile Cluster [J]. IOP Conference Series: Materials Science and Engineering, 2018 (336): 012017.
- 8 Saputra D M, Saputra D, Oswari L D. Effect of Distance Metrics in Determining K - Value in K - Means Clustering Using Elbow and Silhouette Method [C]. Indonesia: Sriwijaya International Conference on Information Technology and Its Applications (SICONIAN 2019), 2020.
- 9 王建仁, 马鑫, 段刚龙. 改进的 K - means 聚类 k 值选择算法 [J]. 计算机工程与应用, 2019, 55 (8): 27-33.