

# 基于多维数据模型的医院信息管理系统研究

王雪, 高姝

(首都医科大学附属北京地坛医院, 北京 100015)

**摘要:** 中国人口变化对医院业务效率提出了新需求, 优化医院信息管理系统是提高医院业务效率最直接的方法, 但目前多数医院信息管理系统不能很好地发挥医院数据的价值。因此研究提出一种融合聚类算法和多维数据模型的医院信息管理系统模型。模型采用多维数据模型的思想, 能更全面地对医院数据进行分析 and 归类。K 均值聚类和模糊聚类的引入进一步增强了数据的聚类分析效果。实验结果显示优化后模型的数据利用率最高达到了 97.49%、分类准确率最高达到了 91.70%。证明优化后的模型在数据利用率和数据分类准确率方面有了明显提升, 因此该优化方案具备可行性。

**关键词:** 医院信息系统; 多维数据模型; 模糊聚类; K-means 聚类

中图分类号: TP391 文献标识码: A 文章编号: 1003-7241(2025)09-0149-05

## Research on Hospital Information Management System Based on Multidimensional Data Model

WANG Xue, GAO Shu

(Beijing Ditan Hospital, Capital Medical University, Beijing 100015, China)

**Abstract:** The demographic changes in China have imposed new demands on the operational efficiency of hospitals. optimizing hospital information Management Systems (HIMS) serves as the most direct approach to enhance the operational efficiency of hospitals. Nevertheless, the majority of existing HIMS fail to fully exploit the value inherent in hospital data. Consequently, this study proposes a novel HIMS model that integrates clustering algorithms with a multi-dimensional data model. The proposed model adopts the methodology of the multi-dimensional data model, which enables a more comprehensive analysis and classification of hospital data. The incorporation of K-means clustering and fuzzy clustering further strengthens the effectiveness of data clustering analysis. Experimental results demonstrate that the optimized model achieves a maximum data utilization rate of 97.49% and a maximum classification accuracy of 91.70%. These findings confirm that the optimized model exhibits significant improvements in both data utilization rate and data classification accuracy, thereby validating the feasibility of the proposed optimization scheme.

**Keywords:** hospital information system; multidimensional data model; fuzzy clustering; K-means clustering

### 0 引言

目前医院信息管理系统涵盖医院的财务、门诊以及住院等业务, 不同业务对应着不同的子系统。然而传统的医院信息管理系统往往只关注数据的简单录入和查询, 无法满足现代医院对数据处理、信息共享和决策支持等方面的需求。在实际生活中大多数医院不同子系统间信息共享程度不高, 数据管理也不够统一。李雅兰等提出一种雾辅助的隐私保护分层多维数据聚合机制, 多维数据可为隐私数据提供不同粒度的数据应用支撑, 达到更好的隐私信息保护<sup>[1]</sup>。郝旭龙等提出一种基于智能电网中隐私保护多维数据聚合方案, 从而实现电网的细粒

度数据的精确分析, 提高了用户匿名性, 机密性和数据完整性<sup>[2]</sup>。Suppa N 等利用多维数据模型构建了多维贫困指数工具箱, 该技术促进了单字长定点指令平均执行速度 (million instructions per second, MIPS) 的规范、预估和分析, 推动了 MIPS 技术的更新和发展<sup>[3]</sup>。随着数据挖掘技术的发展, 医院的众多病例数据对疾病和药物应用的研究有极大的帮助, 但多数医院信息管理系统不具备数据挖掘和聚类分析的功能, 导致大量的数据无法发挥其价值。多维数据模型通过对医院不同类型的数据进行多维度定义, 从而支持数据在多个维度之间关联, 提高了数据间的联系, 同时为数据共享和数据的分析提供便利<sup>[4]</sup>。鉴于上述原因, 此次研究采用多维数据模型、K 均值聚类以及模糊聚类技术, 对医院传统的信息管理模型进行优化, 旨在提高医院信息管理系统对数据的分析处理能力, 促进医院信息化建设。

\*基金项目: 北京市医院管理中心北京市属医院科研培育计划项目 (PG2020001)

收稿日期: 2024-01-04

# 1 融合聚类分析与多维数据模型的医院信息管理系统研究

## 1.1 基于多维数据模型的医院信息管理系统构建

通常医院信息管理系统是一个组合系统,该系统下包含多个子系统,常见的子系统有护理系统、医院财务系统、设备物资系统、血库系统、医务系统等等,根据不同医院开设业务的差异略有不同。常见的医院管理子系统及其功能如表1所示。

表1 医院信息管理子系统及其功能汇总

子系统名称	主要业务	使用科室
门诊挂号系统	预约挂号	门诊部、急诊部
住院管理系统	护士站办公、住院部医生办公	住院部
病案系统	病案编目、病案存储与管理	病案室
检查系统	超声、核磁等检查报告存储与传输	检验科、放射科
血液管理系统	血库及患者检查血液样本的管理	检验科、血液净化科
药品管理系统	记录药品入库与出库信息	中/西药房
收费系统	与医保网相连,用于患者就医收费	收费处

数据仓库作为医院操作数据的媒介,可用于存储、查阅、更改医院内部数据,对医院信息管理系统建设具有重要影响。数据仓库通过源事物系统的副本进行信息维护,可以将多个源的数据聚集在一起统一处理,增强了数据的一致性<sup>[5]</sup>。一个完整的数据仓库体系结构如图1所示。

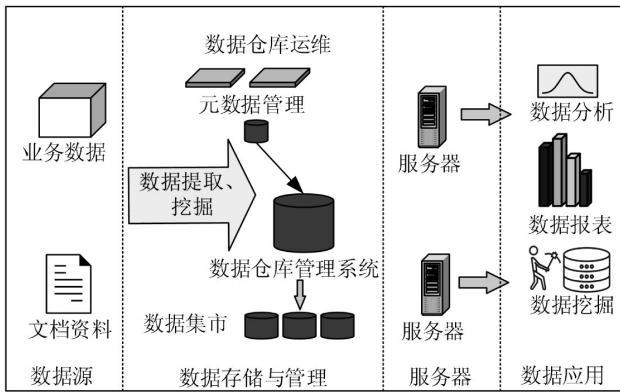


图1 数据仓库结构

数据仓库包括数据源、暂存区、数据提取层、系统业务层、数据逻辑层和数据表示层等<sup>[6]</sup>。数据仓库通常采用自底向上的方法进行构建,即先创建数据集市以及数据维度和事实的业务流程,该流程使不同维度数据之间共享特定的事实,以此增强数据间的联系。多维数据分析模型是对数据仓库的一种改进,其具有易于维护、输出数据直观和高性能等优点,被各大医院信息管理系统建设方案选择。多维数据模型由逻辑数据集、逻辑措施、维度和数据属性等构成,这些组成部分可以用来描述数据的各个方面<sup>[7]</sup>。其中,逻辑数据集用于组织数据,可以对同一数据集中具有相同属性的数据进行分析并显示在一起。逻辑措施用于多维数据收集业务操作,措施通常会

通过批处理进行定期更新。维度这个概念包含一个全局唯一的值,它被用于区分和分类数据。在实际应用中,措施通常是多维度的,为了使一个价值衡量具有意义,它必须由每个维度的成员资格决定。数据属性为当前数据提供了额外的附加特性,附加的特性可以辅助数据的区分和分类。

多维数据关系模型的实现常见有星型模式和雪花模式两种,此次研究选用星型模式实现关系模型的构建。星型模型包含数据维度表、事实表以及汇总表,该模式下,所有数据均在元数据中进行存储,且元数据需要通过一个逻辑函数作为多维对象进而对数据进行处理。

医院信息管理系统多维数据模型构建的基础是数据仓库的设计,数据仓库必须包含医院各部门的所有数据,并为数据赋予相应的维度。数据仓库对数据处理前通常会将数据进行归一化处理,归一化计算如式(1)所示。

$$e = \sum_{i=1}^m p_i \cdot \ln p_i \tag{1}$$

式中, $e$ 为系统熵值; $m$ 表示系统状态数量; $p_i (i=1, 2, \dots, m)$ 表示系统状态概率。

系统熵值是一种衡量系统不确定性的系数,表示系统内部混乱程度或随机性的程度。熵值越大,意味着系统在该状态下的不确定性越大,因此,在计算系统状态权值时,熵值越大,该状态的权值就越小。对于一个系统来说,如果其状态的不确定性较高,那么该状态对于整个系统的贡献度就会降低。相反,如果一个系统的熵值较小,那么其状态的不确定性就相对较低,该状态对于整个系统的贡献度就会相对较高。熵权的计算如式(2)所示。

$$w_e = -k \cdot \sum_{i=1}^m p_{ij} \cdot \ln p_{ij} \tag{2}$$

式中, $k$ 为常数,其表达式为 $k = \frac{1}{\ln^m}$ ,在该表达式中的 $w_e$ 值应处于 $[0, 1]$ 。

多维数据模型下的医院信息管理系统状态由不同维度的系统状态共同决定,因此需要计算出各维度下系统的权值系数矩阵,该计算的数学表达式如式(3)所示。

$$w_j = \frac{d_j}{\sum_{j=1}^n d_j} \tag{3}$$

式中, $d_j$ 表示各子系统的一致度量,其计算表达式为 $d_j = 1 - w_e$ ;  $n$ 表示子系统个数。

在已知每个维度下各子系统的权值系数后就可以对整个系统的不确定性进行评估,该评估通过多维度状态及其对应权值系数进行,评估结果如式(4)所示。

$$S_i = \sum_{j=1}^n w_j \cdot H_{ij} \tag{4}$$

式中, $H_{ij}$ 表示 $j$ 维度下第 $i$ 个子系统状态指数。

系统的整体评估结果将直接影响数据仓库的运行,因此引入多个维度的熵值可以从多个层面反映系统的状态,进而使数据仓库更全面地统合数据。

### 1.2 融合聚类算法的多维数据信息管理系统模型构建

聚类是非监督学习算法的一种,聚类过程则是将具有相似特征的对象集合在一个集群中,常见的聚类算法有K-means、模糊聚类、层次聚类等<sup>[8]</sup>。层次聚类对数据量敏感不适用于数据存储量较大的医院数据,且该聚类算法可能出现空簇,会浪费医院服务器资源<sup>[9]</sup>。因此,此次研究使用K-means聚类算法与模糊聚类算法对多维数据分析模型进行优化。对于高维数据,各个特征向量之间存在一个距离测量,目前常用的测量公式如式(5)所示。

$$D_p(x_i, x_j) = \left( \sum_{k=1}^n |x_{i,k} - x_{j,k}|^p \right)^{\frac{1}{p}} \quad (5)$$

式中,  $x_i = (x_{i1}, x_{i2}, \dots, x_{in})^T$  与  $x_j = (x_{j1}, x_{j2}, \dots, x_{jn})^T$  均为  $n$  维空间中的两个点向量;  $p$  为变参数,当  $p=1$  时,它就是曼哈顿距离;当  $p=2$  时,它就是欧氏距离;当  $p \rightarrow \infty$  时,它就是切比雪夫距离;  $n$  表示数据维度;  $x$  均为数据库中的数据。K-means 聚类算法是最简单的无监督学习算法之一,该方法通过输入的  $K$  值将数据聚合成  $K$  个聚类簇,并计算出每个聚类簇中数据的平均值作为该簇的中心。在K-means聚类算法中目标函数如式(6)所示。

$$M = \sum_{j=1}^k \sum_{i=1}^n \|x_i^{(j)} - c_j\|^2 \quad (6)$$

式中,  $\|x_i^{(j)} - c_j\|^2$  用于测量数据点  $x_i^{(j)}$  到集群中心  $c_j$  的距离;  $k$  是输入值;  $c_j$  是集群中心,也是聚类簇的均值;  $n$  表示数据的维数;  $\|\cdot\|$  表示任意表达之间的相似性测量数据和中心。

K-means 算法的缺陷在于该算法必须提前给出聚类的数量,而医院数据的分类并不是任何情况都是一直分类数的,因此该算法无法达到医院数据的处理需求<sup>[10]</sup>。鉴于上述K-means算法的不足,此次研究引入模糊聚类对K-means算法的不足进行改进。模糊聚类基于最小化目标实现,因此该方法的目标函数如式(7)所示。

$$M_m = \sum_{i=1}^N \sum_{j=1}^C u_{ij}^m \|x_i - c_j\| \quad 1 \leq m < \infty \quad (7)$$

式中,  $N$  表示集群项数;  $C$  表示聚类中心数;  $u_{ij}$  表示第  $j$  个集群中第  $i$  个  $x$  的隶属程度;  $m$  为大于 1 的常数,通常通过随机数函数获取。  $u_{ij}$  的计算表达式如式(8)所示。

$$u_{ij} = \frac{1}{\sum_{k=1}^N \left( \frac{\|x_i - c_j\|}{\|x_i - c_k\|} \right)^{\frac{2}{m-1}}} \quad (8)$$

式中,  $c_k$  表示第  $k$  个集群的聚类中心。聚类中心  $c_j$  的计算公式如式(9)所示。

$$c_j = \frac{\sum_{i=1}^N u_{ij} \cdot x_i}{\sum_{i=1}^N u_{ij}} \quad (9)$$

在K-means聚类中,聚类中心是通过采用模糊聚类方法计算出来的。这种方法可以有效地避免出现空簇的情况,因为模糊聚类能够更好地处理数据中的不确定性。此外,模糊聚类还可以降低算法对数据量变化的敏感性,使得聚类结果更加稳定和可靠。此时数据的聚类过程如图2所示。

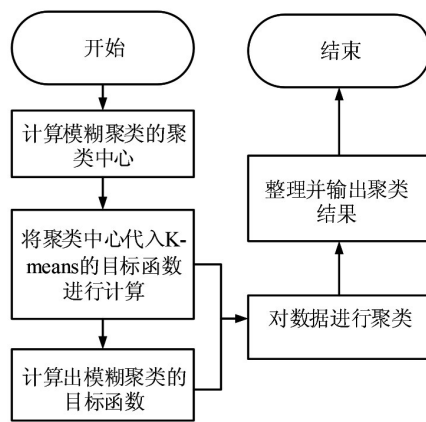


图2 融合聚类计算流程图

模糊聚类处理是一种先进的数据分析方法,它能够处理那些具有不确定性和模糊性的数据<sup>[11]</sup>。与传统的分类方法相比,模糊聚类能够更好地反映实际情况,该算法能够考虑到数据之间的相似性和差异性,避免了“非此即彼”的分类方式,使数据应用和分析具有多样性,从而更加准确和全面地应用医院数据<sup>[12]</sup>。此次研究将K-means算法和模糊聚类算法进行结合,得到的融合算法计算复杂度较低,可同时处理大量数据。由于模糊聚类的不确定性使得融合算法也能处理多模态问题。此时医院信息管理系统的数据仓库构建完成,该数据仓库运行流程如图3所示。

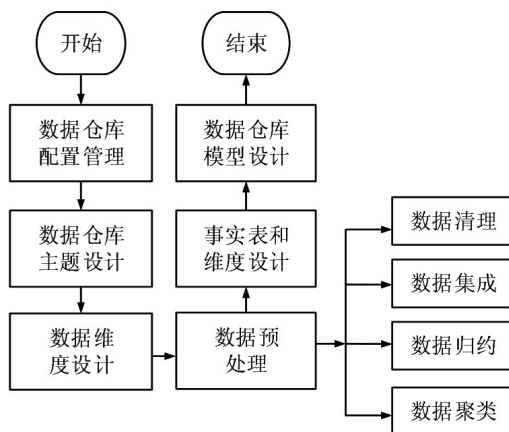


图3 融合聚类算法的数据仓库运行流程

优化后的医院信息管理系统能够提供高效方便的数据分析环境、处理大量数据,并能够集成多种数据,从而提升数据的丰富性和价值。

## 2 基于多维数据模型的医院信息管理系统应用分析

实验环境的设备是一台配备有 Intel Core i7-9750H @ 2.60 GHz 六核中央处理器的高性能台式电脑,拥有 16 GB 的运行内存。系统资源为 Windows 11 操作系统,数据仓库搭建采用 SQL 数据库作为数据存储和处理的核心工具,数据清洗、预处理和分析使用了 Python 3.8 及其编译器。此次研究将时间维度和系统维度对数据共享效率的影响进行对比,系统维度选取医院收费系统(hospital billing system, HBS)、门诊挂号系统(outpatient registration system, ORS)、住院管理系统(hospitalization management system, HMS)和病案系统(medical record system, MRS),时间维度选取 2023 年 1 月~7 月。利用 Python 3.8 及其编译器在创建的数据仓库中进行数据共享,记录模型运行中相关数据的数据共享效率,并绘制为图 4。从图 4 中可以看出,医院数据也具有时效性,近两个月的数据共享效率远高于久远的数据。此外 HMS 和 ORS 的数据共享效率高于 HBS 和 MRS, MRS 的数据共享效率仅 57.11%, 该系统数据共享效率最低原因是病人的病案涉及隐私,因此病案中涉

及患者个人信息的部分数据无法共享。共享效率最高的是 HMS,达到了 78.82%。由于住院患者需要实时与收费系统、药品管理系统等进行数据传输,所以该系统的数据变更频繁,因此数据共享效率更高。

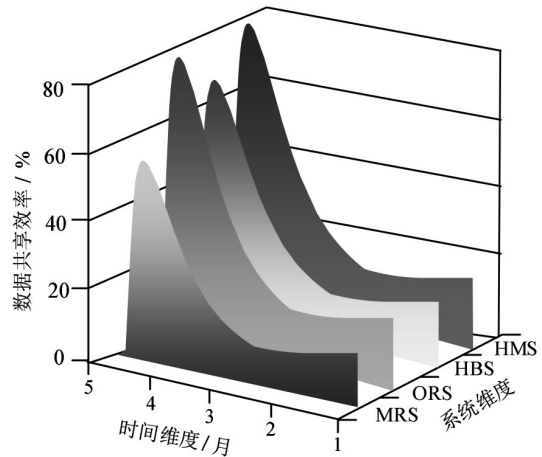
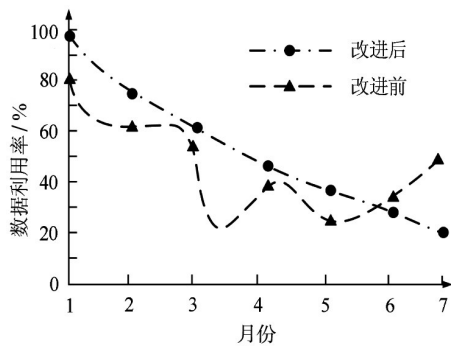
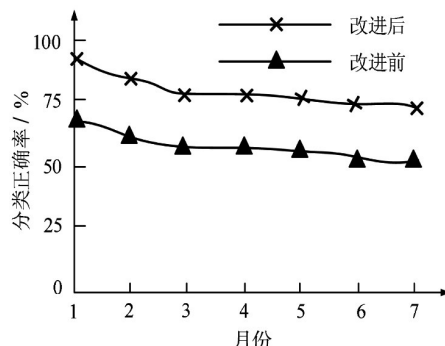


图 4 多维数据分析模型中数据共享效率展示图

为了分析管理系统改进前后医院数据的利用率和分类准确率,研究选取 2023 年 1 月-7 月的数据进行实验,实验结果如图 5 所示。从图 5(a)中可以看出,改进后的系统数据利用率最高可达 97.49%,比改进前有了明显提升;从图 5(b)中可以看出,改进后的系统数据分类准确率最高达到了 91.70%,高于改进前的 70.06%。



(a) 数据利用率对比



(b) 分类准确率对比

图 5 改进前后系统数据利用率和分类准确率对比

多维数据模型能从更高的维度表述数据的特征,对后续数据聚类分析有积极影响。为了分析此次研究提出的系统模型在数据分析和聚类方面的性能,将门诊数据、收费数据、住院数据作为输入,观察输入数据聚类后的输出结果,结果如图 6 所示,使用融合聚类算法前的模型聚类密度低于使用融合聚类算法后的模型。同时使用融合聚类算法后,不同簇之间的链接点变多,说明该聚类方法更好地保留了数据特征的多样性,更好地避免了“非此即彼”的聚类结果。

经过实验分析,此次研究提出的融合聚类算法与多

维数据模型的医院信息管理系统对数据的利用率和分类准确率高于优化前的系统。且改进后,系统的聚类效果有了明显的提高,因此,此次研究对医院信息管理系统改进对医院信息化建设具有积极影响。

## 3 结束语

此次研究旨在增强医院信息系统分析和共享数据的能力,因此提出一个融合聚类分析与多维数据模型的医院信息管理系统数据处理模型。研究中利用多维数据分析模型增强各子系统之间数据共享能力,便于所有数据

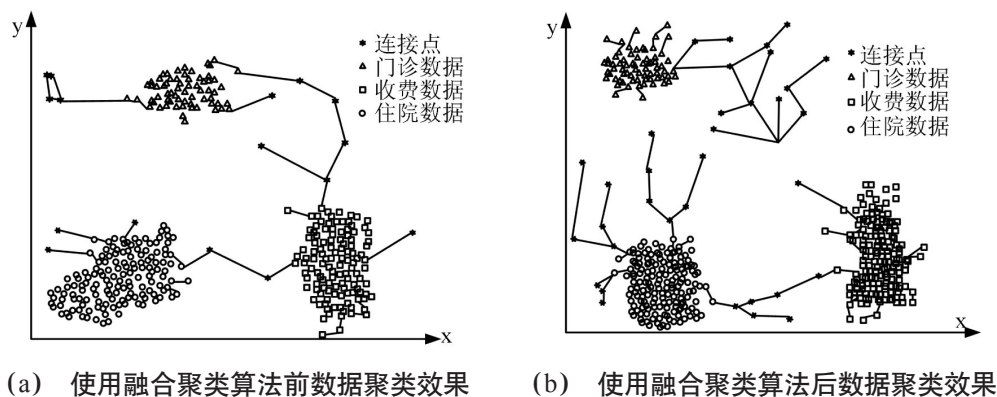


图6 使用融合聚类算法前后聚类效果对比

的统一管理,同时通过结合K-means聚类算法和模糊聚类增强医院信息管理系统对数据的分析和挖掘能力,从而更好地发挥医院数据的价值。经实验得出,优化后的系统对数据的共享能力有了明显提升,在同一数据集上提出模型数据利用率最高达到了97.49%,远高于其他对照模型。此外,提出模型分类准确率最高也达到了91.70%,比改进前提高了21.1%。此外系统的聚类分析能力也相较于优化前得到了提升,相比于优化前,系统聚类密度得到了提升,聚类多样性也得到了充分体现,避免了“非此即彼”的聚类现象。在研究过程中也发现了该优化方案存在不足,如优化后的系统复杂度变高,对设备资源的要求也相应提高,因此该系统建设成本较高,中小型医院无法承担其建设费用。希望后续研究进一步改进,以构建一个更完善的医院信息管理系统。

#### 参考文献:

- [1] 李雅兰,王倩,袁可,等.雾辅助的隐私保护分层多维数据聚合研究[J].小型微型计算机系统,2022,43(7):1499-1504.
- [2] 郝旭龙,刘明曦,董国芳.基于同态加密的智能电网多维数据聚合方案[J].计算机工程与设计,2023,44(4):984-990.
- [3] 李永明.基于数据挖掘技术的心理障碍预测模型[J].微型电脑应用,2021,37(11):161-164.
- [4] 王逸,裴生雷,王煜.基于CSI和K-means-SVR的多指纹库室内定位方法[J].计算机应用,2023,43(5):1636-1640.
- [5] 王梅,宋晓晖,刘勇,等.神经正切核K-Means聚类[J].计算机应用,2022,42(11):3330-3336.
- [6] 黄冬梅,张宁宁,胡安铎,等.考虑多维特征和数据增强的空间负荷预测方法[J].科学技术与工程,2022,22(30):13330-13337.
- [7] TANG Y, PAN Z, PEDRYCZ W, et al. Based kernel fuzzy clustering with weight information granules[J]. IEEE Transactions on Emerging Topics in Computational Intelligence, 2022, 7(2): 342-356.
- [8] 陈翠琴,范亚臣,王林.基于改进Mosaic数据增强和特征融合的Logo检测[J].计算机测量与控制,2022,30(10):188-194.
- [9] 徐昌前,王东,苏峰,等.基于图像数据耦合识别的输电线路安全风险预估方法[J].计算机科学,2023,50(S1):793-798.
- [10] 孙晓.综合性医院数字化档案信息自动分类方法研究[J].自动化技术与应用,2023,42(7):74-77.

[11] HOODA H, VERMA O P. Fuzzy clustering using gravitational search algorithm for brain image segmentation [J]. Multimedia Tools and Applications, 2022, 81(20): 29633-29652.

[12] 刘明群,何鑫,覃日升,等.基于改进K-means聚类k值选择算法的配网电压数据异常检测[J].电力科学与技术学报,2022,37(6):91-99.

作者简介:王雪(1984—),女,本科,研究实习员,研究方向:卫生服务经营管理。