

基于改进 K-Medoids 聚类算法的医院 HRP 系统设计

张蕾¹, 徐叶青²

(1. 南京医科大学第四附属医院, 江苏 南京 210000; 2. 江苏省中医院, 江苏 南京 210029)

摘要:为深入挖掘医院海量信息中有价值的信息,研究面向大型综合三甲医院 C 设计了医院资源规划系统,为推进智慧医院的转型需求,使用 Tent 混沌映射改进人工蜂群优化算法,并在此基础上优化 K-中心点聚类算法。研究提出的算法平均准确率为 93.76%。在实际应用中,2023 年收入预支出预算执行率分别为 99.63% 与 99.04%。上述结果说明研究提出的医院资源规划系统具有优秀的性能,能显著提升实际应用中预算管理,有利于医院财务精细化管理的推进。

关键词:改进 K-Medoids 聚类算法; HRP 系统; 精细化管理; 人工蜂群优化算法

中图分类号: TP311.13; TN197.322

文献标志码: A

文章编号: 1003-7241(2025)12-0142-06

Design of hospital HRP system based on improved K-Medoids clustering algorithm

ZHANG Lei¹, XU Yeqing²

(1. The Fourth Affiliated Hospital of Nanjing Medical University, Nanjing 210000, China;

2. Jiangsu Province Hospital of Chinese Medicine, Nanjing 210029, China)

Abstract: In order to deeply explore valuable information in the massive information of hospitals, a hospital resource planning system is designed for large comprehensive tertiary hospitals C. To promote the transformation needs of smart hospitals, Tent chaotic mapping is used to improve the artificial bee colony optimization algorithm, and on this basis, the K-centroid clustering algorithm is optimized. The average accuracy of the algorithm proposed in the study is 93.76%. In practical applications, the execution rates of the 2023 revenue advance budget are 99.63% and 99.04%, respectively. The above results indicate that the hospital resource planning system proposed in the study has excellent performance and can significantly improve the budget management level in practical applications, which is conducive to the promotion of refined financial management in hospitals.

Keywords: improving K-Medoids clustering algorithm; HRP system; refined management; artificial bee colony optimization algorithm

0 引言

医院资源规划系统(hospital resources planning, HRP)是管理者善用一切手段与资源不断推进医院管理创新的工具,是医院实现管理智能化、精细化与标准化的支撑环境^[1-3]。该系统使得医护分开核算、零库存管理、顺价作价、多方融资以及供应链管理先进管理方法在医院管理中应用^[4-5]。但目前 HRP 系统中存在海量信息资源无法得到有效利用。曹苗苗等提出云模型生物地理学优化算法,同时结合建立医院手术室资源调度模型,结果表明手术等待时间明显降低,对其他医院手术资源调度具有参考价值^[6]。Wider J 为实现精准医疗,为普通人制定疾病治疗与预防策略,提出一种基于基因组数据的云医院系统,以充分考虑不同患者间的差异^[7]。上述研究成果显示了 HRP 系统与精准医疗管理的重要性,但智能运营管理方法日新月异,已难以满足现今的医院财务管理信息化建设。因此研究设计了医院 HRP 系统,并在人工蜂群算法(artificial

bee colony, ABC)的基础上提出基于 Tent 混沌映射(tent chaotic mapping, TCM)的改进算法(improved ABC based on TCM, TIABC),最后对 K-中心点聚类算法进行优化得到基于 TIABC 优化的改进 K-中心点聚类算法(improved K-medoids clustering based on TIABC optimization, TO-K-Medoids)。研究的创新性在于设计了基于 TO-K-Medoids 算法进行数据挖掘的 HRP 系统,并引入财务精细化管理,以促进医院的高质量发展。通过研究以期管理层决策提供潜在有价值的信息,为其他医院的财务智能化建设提供基础支撑。

1 系统构建

1.1 医院资源规划系统的构建

近年来,在大数据技术、人工智能等新技术的不断助推下,医院智能化建设使得医疗机构的管理更加科学化与高效化,以数据为核心资源的数字化时代,正在成为引领与推动新一轮科技革命的核心力量^[8-10]。但目前我国医疗卫生机构的各个业务系统建设与管理相互独立,数据收集需要重复录入,不仅增加了工作量,还造成数据失真等问题,为后续部门协同带来阻碍,同时削弱了资源配置的

* 基金项目:南京医科大学第二附属医院项目(CW201836)

收稿日期:2024-04-12

作用。因此研究以大型综合三甲医院 C 作为研究对象,其包含科研教学、康复保健、医疗与预防的功能。并基于财务精细化管理方法,优化 HRP 系统的设计,积极推动智

能化发展,整合医院的人员、物品与财务信息资源,进而实现医院人员效率的提升以及管理水平的精细化、智能化与标准化。医院资源规划系统的构建架构,如图 1。

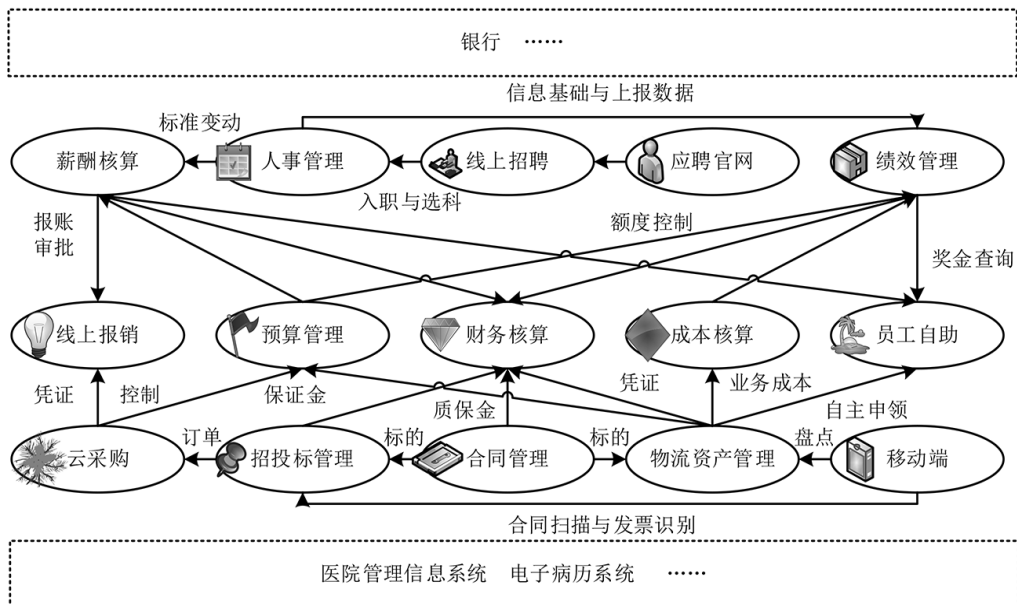


图 1 医院资源规划系统的构建架构

在图 1 中,HRP 系统具有采购管理系统、财务精细化管理系统、人力资源管理系统、医疗业务管理系统等子系统,其中最为关键的子系统是财务精细化管理系统,具有成本核算、财务核算、网上报销、票据管理、决策支撑等模块。而最重要的决策支撑模块是在 HRP 系统利用大数据技术进行分析与数据挖掘的基础上研发的,其有利于固定财务分析任务的减少,医院财务具有众多财务报表,基本上按照一个模板进行制作,通过大数据挖掘技术就可对数据进行智能快速的数据分析,并且在发生异常情况时,还会发出预警,这优化了医院财务风险控制能力。

1.2 基于 Tent 混沌映射的改进 ABC 算法

在 HRP 系统最重要的数据挖掘部分,首先在 ABC 算法的基础上,引入 TCM 进行改进。在 ABC 算法中,通过初始化蜜蜂与设定搜索空间,让蜜蜂在搜索空间中随机选择位置,计算目标函数值,并和其他蜜蜂共享信息,以寻找最优解^[11-13]。ABC 算法中蜜蜂分为观察蜂 g 、采蜜蜂 e 与侦察蜂 z , g 通过 e 提供的信息选择蜜源,并更新蜜源信息; e 与 z 分别负责新的食物源与搜索空间。在初始阶段,每只蜜蜂均是 z , 每个蜜源的个体向量维度以及邻域最大的限制搜索次数分别为 M 与 L , 最大迭代参数为 MAXC, 并通过式(1)随机产生多个蜜源。

$$U_{ij} = U_{\min}^j + \text{rand}(0,1)(U_{\max}^j - U_{\min}^j), j = 1, 2, \dots, M \quad (1)$$

式中, U_{ij} 代表随机产生的可行解空间, U_{\max}^j 与 U_{\min}^j 分别为可搜索解空间的上限与下限, i 代表第 i 只蜜蜂。假设蜜蜂共有 N_i 只, 通过随机蜜源的适应度值 fit_i 进行排序, 就可将 z 进行转换, 得到各 $N_i/2$ 的 e 与 g 。其中 fit_i 的表达式为

$$fit_i = \begin{cases} \frac{1}{1 + f_i}, & f_i \geq 0 \\ |f_i|, & f_i < 0 \end{cases} \quad (2)$$

式中, f_i 表示随机蜜源的适应度值。在蜂种转换完成后, e 就可在蜜源的周围进行邻域搜索, 计算公式为

$$A_{ij} = U_{ij} + \text{rand}() (U_{ij} - U_{mj}), i \neq m, k = 1, 2, \dots, N_e \quad (3)$$

式中, A_{ij} 表示 e 在 U_{ij} 周围搜索的新蜜源, $\text{rand}()$ 的随机搜索范围为 $[-1, 1]$, U_{mj} 代表索引值为 m 时对应的蜜源位置, j 表示 M 中的第 j 维。同时还需评估新旧蜜源的位置, 若新蜜源更好, 就使用新蜜源; 反之就不更换蜜源, 并将对应开采次数加 1。 g 则需在蜂巢进行等待, 通过 e 传递的 fit_i , 根据式(4)得到的概率判断是否跟随。

$$fit_i = \frac{f_i}{\sum_i f_i} \quad (4)$$

蜂群在每次迭代时都进行一次局部搜索与全局搜索, 当任一 e 与 g 在邻域搜索次数 B_i 超过 L 后, 现蜜源的位置仍未更新, 就放弃现位置的采蜜工作, 并且原 e 与 g 转换为 z , 新蜜源位置 0 则通过式(5)获取。

$$U_i(n) = U_{\min} + \text{rand}(0,1)(U_{\max} - U_{\min}), B_i \geq L \quad (5)$$

记录目前全局最优解 Bes, 然后返回到随机新蜜源进行领域搜索, 直到达到 MAXC 或符合停止准则时, 就可输出 Bes。综合上述内容可得 ABC 算法流程, 见图 2。

虽然 ABC 算法具有收敛速度快与求解质量高等优势, 但存在容易陷入局部最优和早熟等缺陷。而 TCM 对应的映射值分布均匀平滑, 因此研究使用其优化 ABC 算

法的蜂群多样性,但这会导致算法后期震荡情况的发生,而通过增设全局最优因子 u_j^g 与交叉操作就可解决上述问

题^[14-15]。TCM 具有更好的复杂性、随机性、灵活性与敏感性依赖,可通过该方法产生初始蜜源,计算公式为

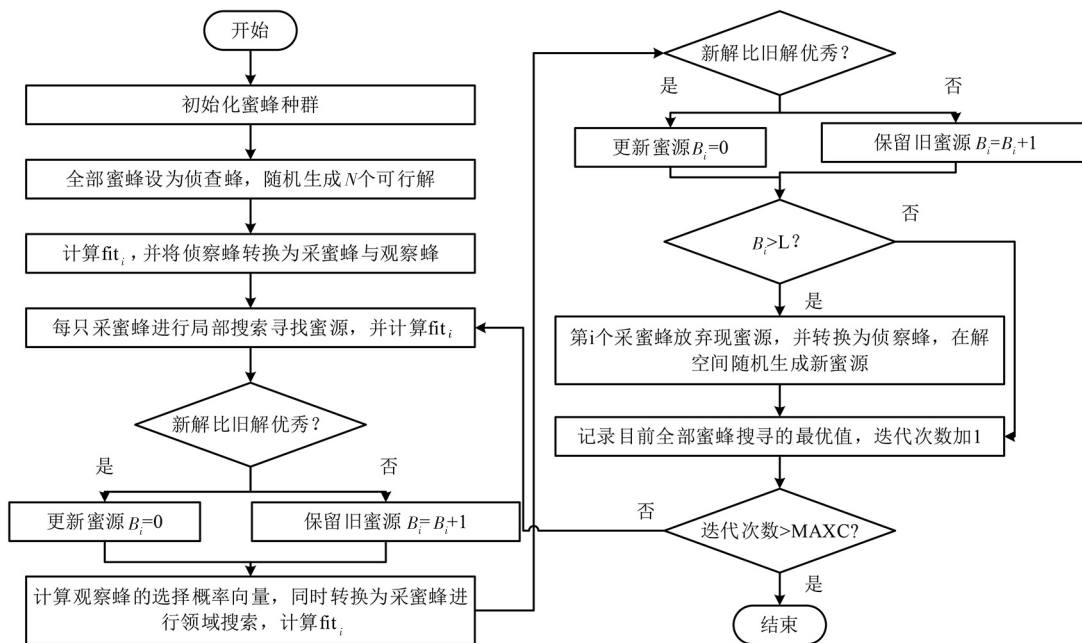


图2 ABC算法流程示意图

$$v_{b,h+1} = \begin{cases} \frac{v_{b,h}}{q}, & 0 \leq v_{b,h} < q, b = 1, 2, \dots, N \\ \frac{1 - u_{b,h}}{1 - q}, & q \leq v_{b,h} \leq 1, h = 1, 2, \dots, d \end{cases} \quad (6)$$

式中, b 与 h 分别为种群规模数目与混沌序号, q 表示混沌参数, 设为 0.7。通过式(6)可优化种群初始化计算公式为

$$U_{ij} = U_{\min}^j + v_{b,h+1}(U_{\max}^j - U_{\min}^j) \quad (7)$$

最后在邻域搜索后,增加 u_j^g 与交叉操作,最后通过贪婪法则确定新解。研究使用二项交叉方法对获取的新解与 u_j^g 进行操作,即

$$v'_{ij} = \begin{cases} v_{ij}, & \text{rand} < \text{cr}, \text{cr} = 0.6 \\ u_j^g + \zeta(u_j^g - v_{ij}), & \text{else} \end{cases} \quad (8)$$

式中, cr 表示平衡算法探索和开发能力的系数, ζ 为交叉系数。

1.3 基于 TIABC 优化的改进 K-Medoids 聚类算法

数字经济时代下,数据已经成为国家基础性战略资源,如何深入挖掘医院 HRP 系统中数据潜在的有价值的信息,是推动医疗信息化建设的重要抓手与核心内容。由于 K-Medoids 聚类算法通过选择实际的观测值作为中心对象,能很好应对数据挖掘中噪声与离群点的情况,具有较好的稳定性。但该聚类算法因为初始聚类中心点的特性,造成最优聚类效果不佳,因此研究使用 TIABC 算法进行优化。设定具有 N 个数据对象的数据集为 $X_i = \{x_1, x_2, \dots, x_N\}$, 每个数据对象包含 S 维特征,现计划将 X_i 分成 k 个类簇 $C_j = \{c_1, c_2, \dots, c_s\}$, x_{ij} 代表第 i 个数据对象

具有的第 j 个特征值, $\text{dis}(x_i, c_j)$ 是用于衡量 x_i 与 c_j 的空间真实距离,计算公式为

$$\text{dis}(x_i, c_j) = \sqrt{\sum_{a=1}^s (x_{ia} - x_{ja})^2} \quad (9)$$

c_j 的质量通过式(10)进行度量。

$$G = \sum_{i=1}^k \sum_{p \in C_j} \text{dis}(p, c_j) \quad (10)$$

式中, p 表示数据集里全部对象。K-Medoids 聚类算法一般使用代价函数评估聚类质量是否优化,计算公式为

$$F = G_2 - G_1 \quad (11)$$

式中, G_1, G_2 分别为替换前数据集、替换中心点数据集对应的全部代表点和类簇中心点误差总和。最后当新旧聚类中心点的 F 值达到阈值或者不再变化时,就可输出最优的簇集合。为提升算法的收敛速度,研究在 K-Medoids 聚类算法的基础上,引进高斯相似度 σ 对目标函数优化,计算公式为

$$G = \sum_{i=1}^k \sum_{p \in C_j} e^{-\frac{\|x_i - c_j\|^2}{2\sigma^2}} \quad (12)$$

此外 K-Medoids 聚类算法由于初始聚类中心的随机性,导致迭代速度慢,因此研究使用 TIABC 算法进行优化,可得 TO-K-Medoids 算法的流程示意图,见图3。

图3中 TO-K-Medoids 算法令 K-Medoids 聚类算法的作为计算适应度值的方法,在 TIABC 算法获得最优解后,将其输入为 TO-K-Medoids 算法的初始聚类中心,然后经过聚类并更新蜂群,直到满足停止条件后就可结束算法。

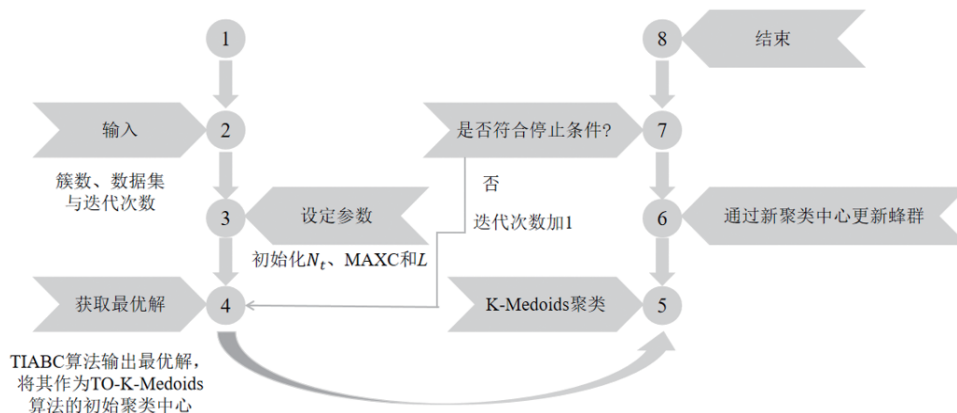


图3 TO-K-Medoids 算法的流程示意图

2 系统实验、应用与分析

研究首先为检验医院 HRP 系统的 TO-K-Medoids 算法的性能,选择目前常用的准确率、运行时间、聚类精度 (clustering accuracy, CA)、调整兰德系数 (adjusting the randall coefficient, ARC) 与加速比 (speed up, SU) 五种指标进行评价。实验环境为 Window 10 操作系统,8 GB 内存与 1 TBG 硬盘的电脑,软件为 MATLAB2022。参数设定如下, N_t 、 L 与 MAXC 分别为 20、200 与 20。数据集选择主流的公共数据集,即 university of california, irvine (UCI) 数据集。选择该数据集中的五个数据集进行性能验证,包含 Iris 数据集 (A)、Glass 数据集 (B)、CMC 数据集 (C)、Wdbc 数据集 (D) 以及 Segmentation 数据集 (E),具体内容见表 1。

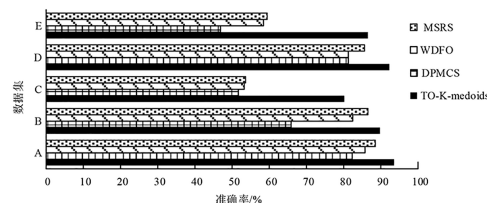
表 1 实验数据集的具体内容

数据集	对象数量/个	k	S	样本对象说明
A	150	3	4	鸢尾属植物的花瓣、花萼长和宽
B	214	6	9	不同类型玻璃的成分信息
C	1473	3	9	不同避孕方式的信息
D	569	2	32	乳腺癌病例编号与肿瘤类别
E	2310	7	19	不同室外图像的信息

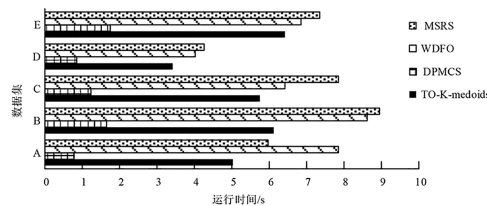
为更科学验证研究提出的算法性能,研究选择目前的主流数据挖掘算法进行对比实验,即基于动态编程匹配的聚类序列算法 (cluster sequence algorithm based on dynamic programming matching, DPMCS)、基于加权深度森林的离群数据挖掘算法 (outlier data mining in weighted deep forest, WDFO) 以及基于最小简约的粗糙集数据挖掘算法 (rough set data mining based on minimum simplicity, MSRS)。研究对不同算法分别进行 50 次实验,并计算准确率与运行时间均值进行对比。

图 4(a) 与图 4(b) 分别为不同算法的聚类准确率与运行时间结果,从图 4 可以观察到,在数据集 A、B 与 D 中四种算法的准确率均超过 80%,同时 TO-K-Medoids 算法的准确率最高,平均为 93.65%;在数据集 C 中,四种算法的

聚类准确率最差,仅 TO-K-Medoids 算法准确率为 80.17%,其余算法的准确率均不超过 60%。在具有规模较大数据量的数据 D 与 E 中,TO-K-Medoids 算法仍能维持较高的准确率,分别为 92.98% 与 86.52%。在运行时间结果对比中,仅 DPMCS 算法的运行时间最快,其余算法均表现出较长的运行时间,这可能是其余三种算法中引进了智能改进算法,导致运行时间增长。虽然 TO-K-Medoids 算法运行时间较长,平均为 5.23 s,但仍比 WDFO 算法与 MSRS 算法的运行时间要少。为进一步验证研究提出的算法的并行化处理能力,研究基于 MapReduce 框架,对不同算法进行实验,同时分析算法在大规模数据集群的运算能力。



(a) 准确率

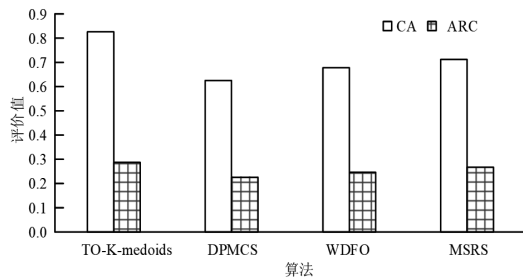


(b) 运行时间

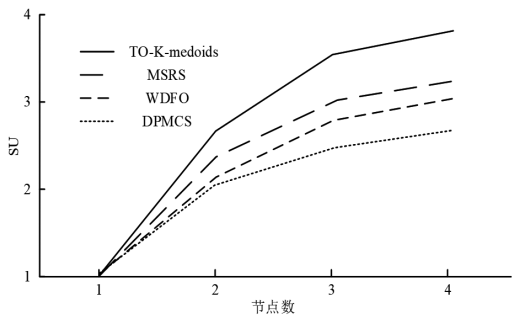
图 4 UCI 数据集下不同算法的性能结果

图 5(a) 与图 5(b) 分别为不同算法的 CA 与 ARC 结果,以及 SU 结果。图 5 显示,DPMCS 算法的 CA 与 ARC 均是最低值,这说明该算法对于具有大数据量与高维数据的情况聚类效果非常差。而 TO-K-Medoids 算法的 CA 与 ARC 值均是最高的,分别为 0.826 与 0.287。在 SU 结果中,TO-K-Medoids 算法仍然是最高的,平均 SU 值为 3.26。综上所述,研究提出的 TO-K-Medoids 算法具有优秀的性能,并且在大规模数据集群情况下,仍能保持较好的聚类

效果与稳定性。为探究基于 TO-K-Medoids 算法的 HRP 系统的应用效果以及提出财务精细化管理措施,研究从 2020 年-2023 年在大型公立医院 C 进行应用实验。

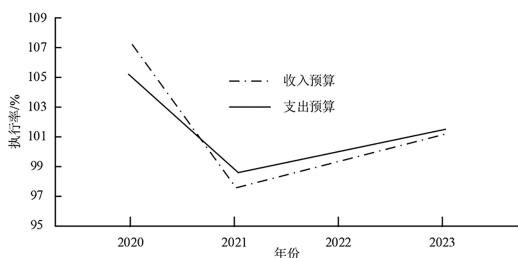


(a) CA 与 ARC 结果

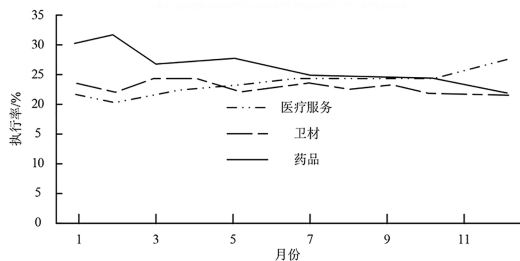


(b) SU 结果

图 5 不同算法的并行化实验结果



(a) 医院 C2020 年-2023 年预算执行率变化曲线



(b) 医院 C2023 年度重要考核指标变化曲线

图 6 基于 TO-K-Medoids 算法的 HRP 系统在医院 C 的应用效果

图 6(a) 与图 6(b) 分别为基于 TO-K-Medoids 算法的 HRP 系统在医院 C 的 2020 年-2023 年预算执行率变化曲线以及 2023 年重要考核指标变化曲线。由图 6 可知,2020 年的收入预支出预算执行率分别为 108.13% 与 105.26%, 2021 年的收入预支出预算执行率分别为 99.21% 与 95.48%; 2023 年收入预支出预算执行率分别为 99.63% 与 99.04%, 这表明在使用基于 TO-K-Medoids 算法的 HRP 系统后,医院 C

的预算管理水平和得到明显提升。和 2022 年同期相比,2023 年的科室药品费用与卫材费用均得到明显的下降,分别下降了 44.76% 与 26.53%, 医疗服务占比则提升到 26.73%, 上述结果表明应用基于 TO-K-Medoids 算法的 HRP 系统后,药品与卫材占比平稳降低,医疗服务收入占比稳定增长。根据上述 HRP 系统的性能与应用分析,可提升财务精细化管理措施。首先是设定科学合理的目标,不同医院因为所处地理环境与经济实力的差异,总体建设目标也存在极大的差异,因此医院需要综合考虑实际现状、地理环境等因素,设定总体战略目标与阶段性目标。然后是医院需要及时优化运营规章制度,充分利用 HRP 系统,推动医院的管理模式,减轻相关人员的工作负担,为居民提供更好的服务,实现医院的高质量发展。最后是医院的经济管理尚未发挥潜力,其也是 HRP 系统建设需要完成的目标,因此可预先对资金投入、物力与人力进行统筹规划,各部门加强协调配合,保障医院财务精细化管理的稳定运行。

3 结束语

针对 HRP 系统中海量数据存在冗余等问题,研究首先基于财务精细化管理方法与大数据技术构建医院 HRP 系统,然后使用 TIABC 算法对 K-Medoids 聚类算法进行优化,最终得到 TO-K-Medoids 算法。实验结果显示,在数据集 A、B 与 D 中四种算法的准确率均超过 80%,同时 TO-K-Medoids 算法在五种数据集中准确率均是最高的。在运行时间结果对比中,仅 DPMCS 算法的运行时间最快,TO-K-Medoids 算法运行时间较长,平均为 5.23 s。并行化实验中,TO-K-Medoids 算法的 CA、ARC 与 SU 值均是最高的,分别为 0.826、0.287 与 3.26。在实际应用中,2023 年收入预支出预算执行率分别为 99.63% 与 99.04%, 预算管理水平得到明显提升。综上所述,研究提出的算法具有优秀的性能,基于此构建的 HRP 系统为医院管理层决策提供有力的数据基础,有效推动了医院财务精细化管理。但研究仍存在不足,研究仅对大型公立医院的应用效果进行检验,该方法在私立医院或者小型公立医院的应用效果尚未进行测试,因此在未来研究中可考虑医院各部门业务结构、医院规模等因素,对医院进行更全面的管

参考文献

- [1] 李佳阳, 李昊, 闫妮, 等. 面向调控云平台的信息化智能搜索技术[J]. 微型电脑应用, 2023, 39(8):32-35.
- [2] 李观义, 肖坚, 陈府玉, 等. 星型灌区监测无线传感器网络 MAC 协议研究[J]. 微型电脑应用, 2023, 39(9):7-10.
- [3] 梅竹松, 严荣国, 徐涛, 等. 多波长红外尿糖检测技术研究[J]. 计算机测量与控制, 2023, 31(5):101-107.
- [4] 刘阳. 基于实战化安全运营的智慧医院网络安全保护体系构建与应用[J]. 中国医疗设备, 2023, 38(11):127-132.
- [5] 郝夏毅, 金顺成, 李崇伟. 基于物联网的智能医用织物管理系统研究与设计[J]. 自动化与仪表, 2022, 37(8):105-107.

(下转第 188 页)