

基于 J2EE 及 K-means 聚类算法的医院绩效考核研究

徐琳

(山西省肿瘤医院 运营管理部,山西 太原 030013)

摘要:研究针对卫生健康领域预算绩效管理的不足,构建了一个基于 J2EE 平台和改进 K-means 聚类算法的医院绩效考核系统。该系统通过数据准备、光电阅读机输入、模糊综合评价、报表处理与系统维护等模块,实现了对医院员工绩效的全面评估。实验结果显示,当该系统的 K 减少到 4 时,ARI 指标有了显著的提升,达到了 0.71 至 0.88 的范围,显示出聚类结果与真实类别的相似度较高。研究提出的改进 K 聚类算法在 XB 值指标上展现出了最佳性能,其 XB 值下降得最为平稳且保持在较低水平。该系统能有效提升医院绩效管理的效率和准确性,为医院管理提供了有力的决策支持。

关键词:J2EE;K-means 聚类;绩效考核系统;模糊综合评价;医院管理系统;聚类

中图分类号: TP311.13;TN234.3

文献标志码: A

文章编号: 1003-7241(2025)12-0177-06

Research on hospital performance evaluation based on
J2EE and K-means clustering algorithm

XU Lin

(Shanxi Cancer Hospital, Operation Management Department, Taiyuan 030013, China)

Abstract: This study addresses the shortcomings of budget performance management in the field of healthcare and constructs a hospital performance evaluation system based on the J2EE platform and improved K-means clustering algorithm. The system achieves comprehensive evaluation of hospital employee performance through modules such as data preparation, photoelectric reader input, fuzzy comprehensive evaluation, report processing, and system maintenance. The experimental results show that when the K of the system is reduced to 4, the ARI index has significantly improved, reaching the range of 0.71 to 0.88, indicating a high similarity between the clustering results and the real categories. The proposed improved K-clustering algorithm demonstrates the best performance on the XB value index, with the XB value decreasing most steadily and remaining at a relatively low level. This system can effectively improve the efficiency and accuracy of hospital performance management, providing strong decision-making support for hospital management.

Keywords: J2EE; K-means clustering; performance evaluation system; fuzzy comprehensive evaluation; hospital management system; clustering

0 引言

目前,预算绩效管理在卫生健康领域仍处于起步阶段,部分卫生健康单位对于预算绩效管理的内涵缺乏深刻认识,行业也尚未形成较为标准的体系和规范^[1-2]。开展医院绩效考核研究显得尤为重要,因此越来越多的研究者开始关注医院绩效系统的建立^[3-4]。李媛媛针对财政预算绩效管理效果评估水平提升的问题,提出基于区块链技术的财政预算绩效管理效果评估模型。该方法通过样本回归分析、模糊综合决策及区块链分布模型,实现效果评估的优化。结果显示,模型可靠性高,评估准确性好,为提升管理能力提供参考^[5]。Fitzpatrick E K R 等为了提高医院的医疗保健提供商和系统的消费者评价(hospital consumer assessment of healthcare providers and systems,HCAHPS)评分的问题,提出考察护理资源(人员配置、技能组合、护士教育和

工作环境)与评分关联性的方法。结果显示,更优的人员配置、高比例的本科学历护士和更好的工作环境与更高的 HCAHPS 评分相关,其中工作环境影响最大^[6]。但这些方法往往侧重于科室或医院的整体绩效,而忽视了医生、护士等个体之间的差异性^[7]。此外,牟杨等提出的基于遗传算法的医务人员绩效评价系统中的遗传算法虽然具有全局搜索能力,但在处理大规模数据时可能效率较低^[8]。而 K-means 聚类算法在处理大规模数据集时具有较高的效率,尤其在经过改进后,能够更好地适应医院绩效考核数据的特点。鉴于此研究融合了数据准备、光电阅读机输入、模糊综合评价、报表处理与系统维护等多个模块,构建了一个全面的医院绩效考核系统。其中,模糊综合评价模块结合了模糊逻辑与 K-means 聚类算法,并针对 K-means 算法的局限性,提出了一种改进的 K-means 聚类算法。期望这种综合性的评价方法和改进的聚类算法,共同为医院绩效考核提供了更为科学、客观的工具,具有显著的新颖性和实用性。

* 基金项目:山西省科技厅战略研究专项(202304031401143)

收稿日期:2024-05-29

1 基于 J2EE 及改进 K 聚类算法的绩效考核平台

1.1 基于 J2EE 平台的医院绩效考核系统设计

在构建基于 Java 2 平台企业版(Java 2 Platform Enterprise Edition, J2EE)平台的医院绩效考核系统中,数据准备模块是基础且关键的一环^[9-10]。该模块通过开发一个用户友好的数据导入界面,使用户能够方便地从医院信息系统等外部数据源导入与绩效考核相关的数据。数据预处理功能采用抽取-转换-加载工具(extract-transform-load, ETL),如 Apache NiFi 或 Talend,自动执行数据的清洗、转换和格式化,为后续的 K-means 聚类算法提供标准化的输入数据^[11-13]。接下来,光电阅读机输入模块通过集成光电阅读机,使得纸质

问卷或报告可以通过扫描识别绩效考核数据^[14-15]。数据解析功能,这需要调用 Java 的 javax. imageio. ImageIO 用于读取图像文件,以及 org. apache. pdfbox. pdmodel. PDDocument 用于处理 PDF 文档,将扫描得到的图像信息转换为系统可处理的结构化数据。在模糊综合评价模块中,通过 J2EE 服务器端实现模糊逻辑和 K-means 聚类算法,构建模糊综合评价模型。该模型作为系统的核心,通过应用数据评价功能,对员工绩效进行综合评价并生成评价结果。相关的 Java 接口可能包括 java. util. List 来存储和处理数据集,以及自定义的算法接口如 FuzzyEvaluator 和 K-meansClusterer,如图 1 所示。

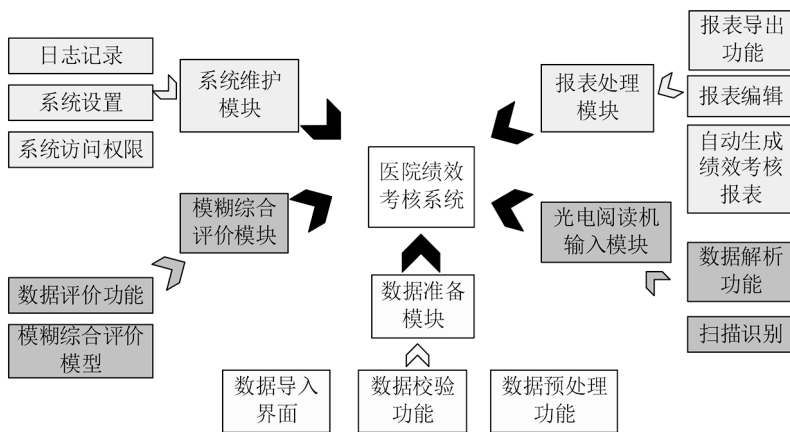


图 1 系统功能模块结构示意图

报表处理模块则利用 J2EE 的报表服务,如 JasperReports,自动生成绩效考核报表。通过调用 org. apache. poi 操作 Office 文档,用户可以通过提供报表编辑和导出功能,定制报表格式,并支持导出到 PDF、Excel 等格式。最后,系统维护模块在 J2EE 应用中实现用户管理,确保不同角色

的用户具有相应的系统访问权限。系统设置功能允许管理员调整系统参数,以符合医院的考核标准和流程。日志记录功能则通过 Java 的日志框架,如 java. util. logging,来监控系统运行状态和进行故障排查,确保系统的稳定性和可维护性。系统的 E-R 结构示意图如图 2 所示。

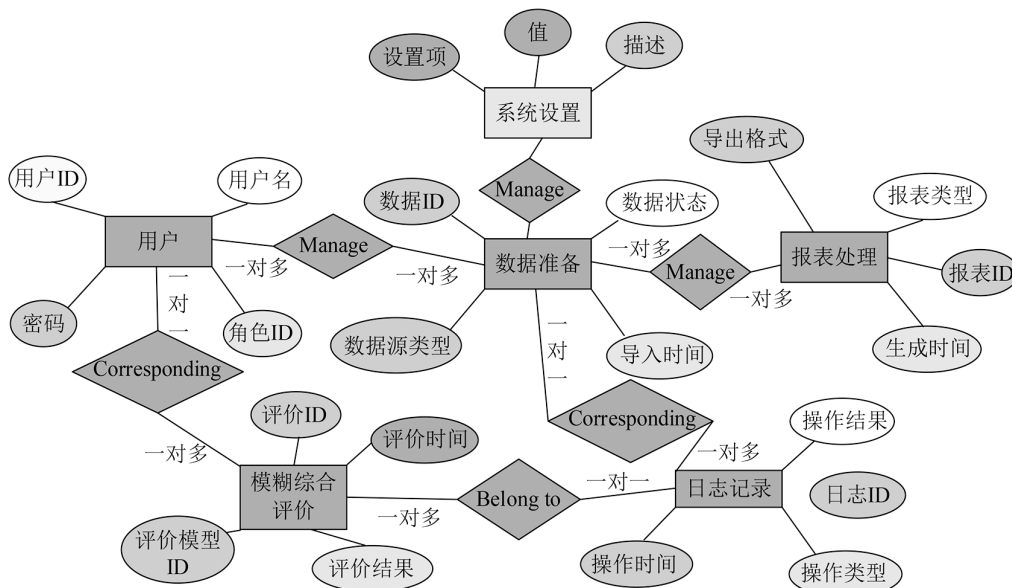


图 2 医院绩效考核系统 E-R 结构示意图

医院绩效考核系统的 E-R 结构示意图清晰地展示了各个核心实体及其关键属性。其中,用户实体包含了用户 ID、用户名、密码和角色 ID 等属性,确保了用户信息的完整性和安全性。而角色实体则通过角色 ID、角色名称和权限描述来定义不同用户的访问权限。在数据流程中,数据准备是关键一环,其属性包括数据 ID、数据源类型、导入时间和数据状态,有效追踪数据的来源和状态。经过模糊综合评价后,评价结果会被记录,包括评价 ID、评价时间等关键信息。同时,报表处理会生成各类报表,详细记录报表 ID、类型、生成时间和导出格式。所有的操作都会被日志记录所捕获,包括操作时间、类型和结果,便于后续的追踪和审计。

1.2 基于改进 K 聚类算法的模糊综合评价模块

在模糊综合评价模块中,通过 J2EE 服务器端实现模

糊逻辑和 K-means 聚类算法,构建模糊综合评价模型。模型的第一步是初始化 K-means 聚类算法,选择初始聚类中心,然后计算每个员工绩效数据点到各个聚类中心的距离,使用曼哈顿距离等度量标准,计算每个员工的绩效数据点与每个聚类中心之间的距离。第三步是分配到最近的聚类,根据计算出的距离,将每个员工分配到最近的聚类中心,从而形成初始的 K 个聚类。第四步是更新聚类中心为分配到该聚类的所有数据点的均值,对每个聚类内的所有员工绩效数据点求均值,以此更新该聚类的中心。第五步是对每个聚类内的数据应用模糊综合评价流程,在每个聚类内部,应用模糊逻辑对员工的绩效进行综合评价。这涉及构建评价指标体系和模糊规则库。模糊综合评价模块算法流程如图 3 所示。

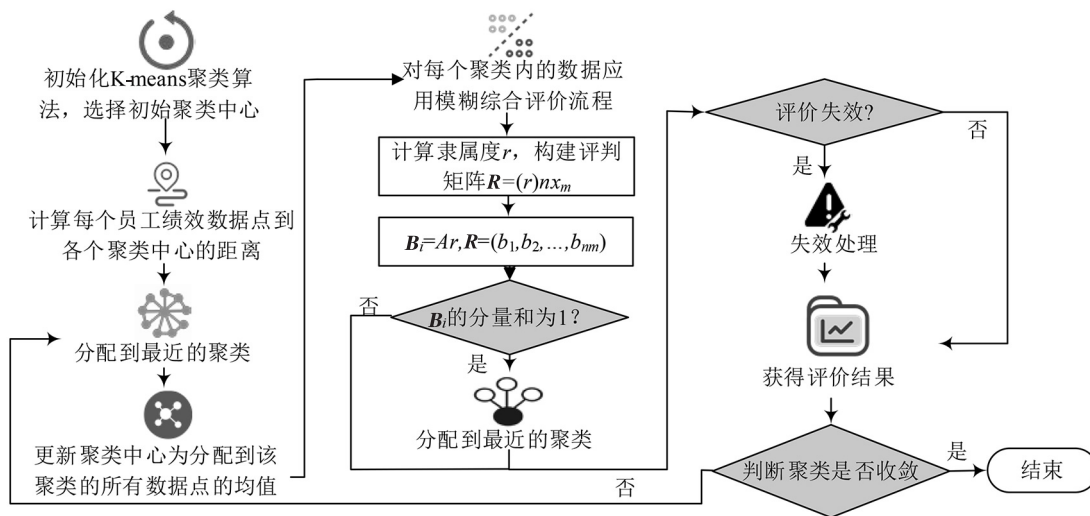


图 3 模糊综合评价模块算法流程

第六步计算隶属度 r ,对于每个评价指标,构建一个 $n \times m$ 的评判矩阵 R ,为

$$r_{ij} = \exp\left(-\frac{(x_{ij} - c_j)^2}{2\sigma_j^2}\right) \quad (1)$$

式中, r_{ij} 是第 i 个员工在第 j 个评价指标上的隶属度。 x_{ij} 是第 i 个员工在第 j 个评价指标上的绩效数据。 c_j 是第 j 个评价指标的理想值或中心点。 σ_j^2 是与第 j 个评价指标相关的标准差。第九步使用隶属度 r 和评判矩阵 R 进行模糊综合评价,得到每个员工的绩效评价向量 B_i ,为

$$B_i = \frac{\sum_{j=1}^m a_j \cdot r_{ij}}{\sum_{j=1}^m a_j} \quad (2)$$

式中, B_i 是第 i 个员工的绩效评价向量。 a_j 是第 j 个评价指标的权重。 r_{ij} 是第 i 个员工在第 j 个评价指标上的隶属度。然后判断 B_i 的分量和为 1,如果是则判断评价失效。检查每个 B_i 向量的分量和是否为 1,如果不是,则说明评

价过程中存在错误,需要重新进行评价。第十步进行归一化处理 B_i ,如果 B_i 的分量和为 1,则对 B_i 进行归一化处理,确保每个分量都是有效的评价值。然后判断聚类是否收敛,检查聚类中心是否稳定,即连续几次迭代后聚类中心的位置变化是否在可接受的范围内,计算连续两次迭代中聚类中心的变化量,并与一个预设的阈值进行比较,其收敛准则为

$$\text{Convergence} = \sqrt{\sum_{k=1}^K \sum_{d=1}^D (C_{kd}^{(t)} - C_{kd}^{(t-1)})^2} < \epsilon \quad (3)$$

式中, $C_{kd}^{(t)}$ 是在第 t 次迭代中第 k 个聚类的第 d 维坐标上的聚类中心值。 D 是数据的维度(即绩效指标的数量)。 k 是聚类的数量, ϵ 是一个预设的小正数阈值,用于判断聚类是否收敛。K-means 聚类算法虽然简单高效,但也存在一些局限性,如对初始聚类中心敏感、容易陷入局部最优、不能很好地处理非球形数据等。鉴于此,研究提出了一种改进 K 聚算法,其算法流程如图 4 所示。

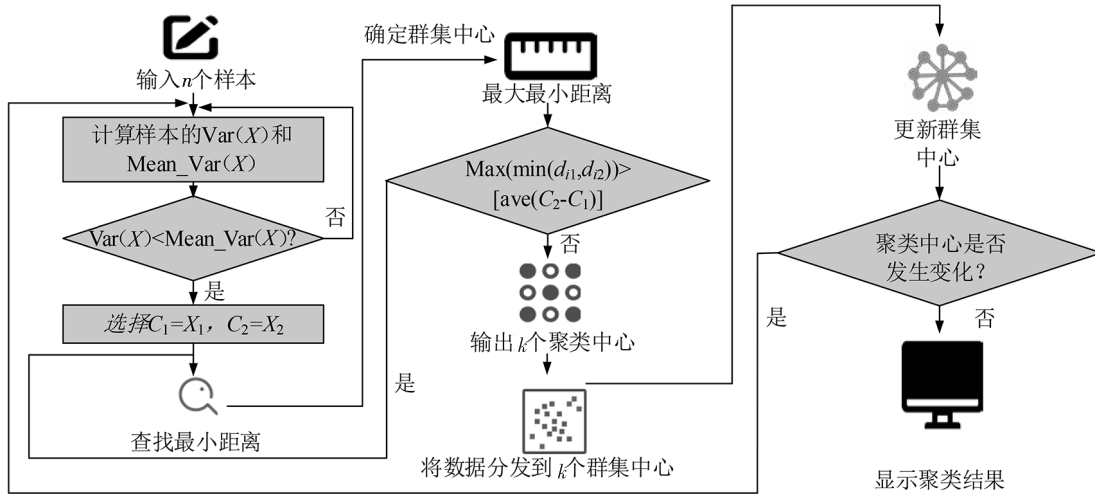


图4 改进K聚类算法流程

该改进的K聚类算法从计算样本的方差 $Var(X)$ 和平均方差 $Mean_Var(X)$ 开始,其计算公式为

$$Var(X) = \frac{1}{n-1} \sum_{i=1}^n (x_i - \bar{x})^2 \quad (4)$$

式中, n 是样本数量, x_i 是每个样本的值, \bar{x} 是样本均值。若 $Var(X)$ 小于 $Mean_Var(X)$, 则直接选择 $C_1 = X_1, C_2 = X_2$ 作为初始聚类中心。反之,则需要查找样本间的最小距离,以此确定初始的群集中心,其中 $Mean_Var(X)$ 为

$$Mean_Var(X) = \frac{1}{n} \sum_{i=1}^n Var(x_i) \quad (5)$$

接着,通过计算样本间的最大最小距离,判断 $Max(\min(d_{i1}, d_{i2}))$ 是否大于群集中心间的平均距离 $ave(C_2 - C_1)$, 这个过程涉及比较样本间最小距离的最大值与聚类中心间平均距离的关系为

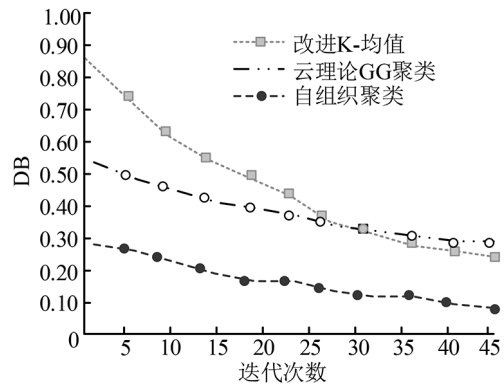
$$Max(\min(d_{i1}, d_{i2})) > ave(C_2 - C_1) \quad (6)$$

式中, d_{i1} 和 d_{i2} 是样本 i 到两个候选聚类中心的距离, $ave(C_2 - C_1)$ 是两个聚类中心间的平均距离。以确保初始群集中心的选择既不过于紧密也不过于分散。一旦确定了 k 个聚类中心,算法便将数据分发到这些中心,并不断更新群集中心,直至聚类中心不再发生变化。

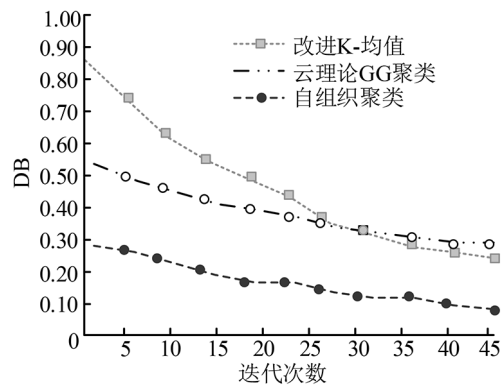
2 医院绩效考核平台应用性实验与性能测试

2.1 模糊综合评价模块性能测试

为了提高医院绩效考核系统的准确性和效率,研究改进了K聚类算法,并基于这一创新算法,设计了一种全新的医院绩效考核系统。为了对该算法的优越性进行验证,研究引入了文献[14]提出的自组织聚类、文献[16]提出的云理论GG聚类以及文献[17]提出的密度聚类算法与研究提出的改进k聚类算法进行对比。在Windows 11平台上利用Python语言实现了上述算法,选择谢贝尼指数(xie beni, XB)和戴维森堡丁指数(davies-bouldin index, DB)作为评价1指标,实验结果如图5所示。



(a) DB



(b) XB

图5 不同聚类模型评价指标统计结果对比

根据图5(a)的展示,可以清晰地观察到,随着迭代次数的递增,三种不同的聚类模型——自组织聚类、云理论GG聚类以及我们研究提出的改进K聚类算法,它们的DB值都展现出了下降的趋势。但值得注意的是,自组织聚类的DB值在整个过程中始终处于最低水平,其最小值甚至达到了0.08,这显示了其出色的聚类效果。相对而言,研究提出的改进K聚类算法虽然也呈现出下降的DB值,但整体水平却相对较高。从图5(b)中可以看出,不同

聚类方法的XB值同样随着迭代次数的增加而逐渐下降。但在这里,自组织聚类的XB值下降得最为显著,云理论GG聚类紧随其后,而研究提出的改进K聚类算法在这一指标上展现出了最佳性能,其XB值下降得最为平稳且保持在较低水平。此外,研究还对比了改进K聚类前后对系统的聚类效果的影响,并将其进行了可视化,实验结果如图6所示。

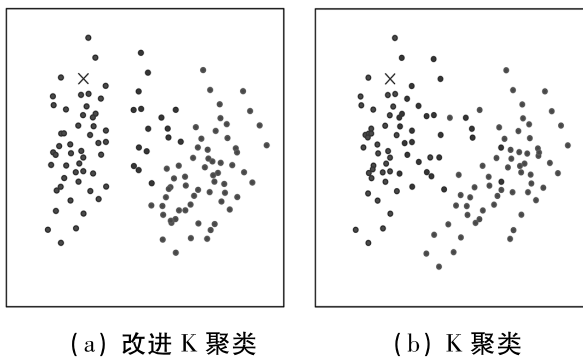


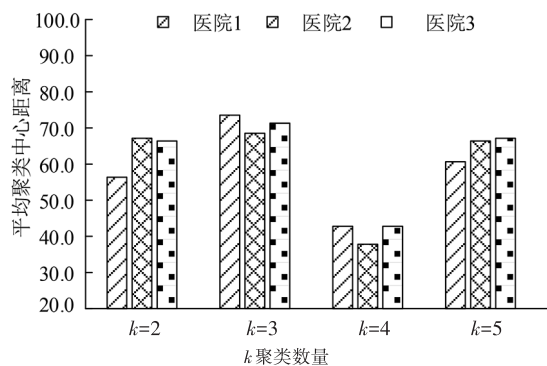
图6 改进K聚类前后聚类效果

从图6(b)中可以明显观察到,未进行改进的k聚类方法在处理数据时显得力不从心,数据点的分类边界模糊不清,这表明分类效果并不理想。同时,还可以看到,每个簇中的数据点在特征空间上分布得较为分散,彼此间的距离较远,这无疑降低了聚类的准确性。然而,在图6(a)中,情况得到了显著的改善。通过模型的优化,它现在能够精准地将高度相似的数据点归为一类,形成了紧密且明显的簇群。

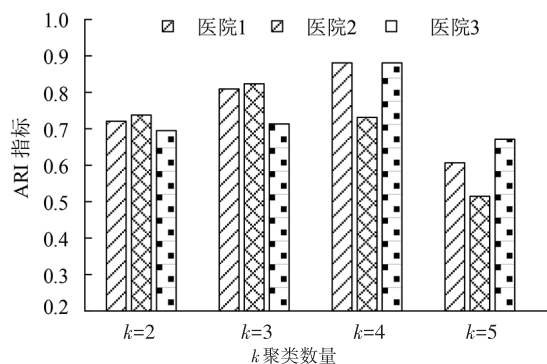
2.2 医院绩效考核平台应用性实验

实验充分验证了研究所提算法的优越性,为了验证该算法在实际的应用中也有同样优异的表现,研究在上海市三所医院中应用了该系统,并探究了ARI指标与评价聚类中心距离随k聚类数量的变化关系,实验结果如图7所示。

根据图7(a)的展示,可以清晰地观察到,在k值为3的时候,各个时期的平均聚类中心距离大致维持在68至75的范围内。而当k值增加到4时,平均聚类中心距离明显减少,落在了38至43的区间内,这表明聚类中心间的距离在k=4时更为紧凑。从图7(b)中可以看出,当k设定为5时,ARI指标的范围在0.51至0.68之间,这意味着聚类结果与真实类别的吻合度处于中等水平。然而,当k减少到4时,ARI指标有了显著的提升,达到了0.71至0.88的范围,显示出聚类结果与真实类别的相似度较高。这一对比明显表明,在k=4时,聚类的效果更佳,更接近于真实的类别划分。在本次医院绩效考核研究中,基于J2EE平台及K-means聚类算法,针对医疗技术、患者满意度、工作量及应急处理能力四项指标,对骨科、皮肤科、耳鼻喉科、内科和神经科五个实验组进行了全面评估,实验结果如表1所示。



(a) 平均聚类中心距离随k聚类数量的变化关系



(b) ARI指标随k聚类数量的变化关系

图7 ARI指标与评价聚类中心距离随k聚类数量的变化

表1 某医院各科室综合绩效评价结果

实验组	医疗技术	患者满意度	工作量	应急处理能力	绩效评价向量
组1	90	85	75	90	[0.85,0.9,0.6,0.85]
组2	70	85	40	90	[0.65,0.9,0.3,0.85]
组3	95	75	80	70	[0.9,0.8,0.7,0.65]
组4	60	90	30	85	[0.55,0.9,0.2,0.8]
组5	75	80	60	75	[0.7,0.8,0.5,0.7]

根据表格1的数据,可以看出五个实验组在医疗技术、患者满意度、工作量和应急处理能力这四项指标上的表现各有差异。组1和组3在医疗技术方面表现较好,而组4在患者满意度方面得分最高。然而,组4的工作量评分最低,可能与科室特性或患者流量有关。应急处理能力方面,组1、组2和组4表现突出。从绩效评价向量来看,各组在各项指标上的相对表现也有所不同,反映了各科室在不同方面的优势和劣势。

3 结束语

本研究通过改进K-means聚类算法,成功构建了一个医院绩效考核系统,并在实际应用中取得了显著效果。在对比实验中,虽然自组织聚类在DB指数上表现最佳,但改进K-means聚类在XB指数上展现了更平稳且低水平的性能,表明其在保持聚类稳定性方面具有优势。在上海市三所医院的应用中,当k值设定为4时,ARI指标显著提升至0.71至0.88,表明聚类效果与真实类别高度吻

合。通过综合绩效评价,各组在医疗技术、患者满意度、工作量和应急处理能力方面表现各异,反映了不同科室的优势与劣势。此系统的应用为医院绩效考核提供了科学、客观的工具,有效提升了管理效率。综上所述,本研究改进的 K-means 聚类算法在医院绩效考核系统中表现优异,为医院管理提供了有力的数据支持。尽管实验结果显著,但系统对数据的预处理和特征选择仍有优化空间。

参考文献

- [1] 陆怡,王鹏,汪卫. 基于子序列相似性的时间序列语义挖掘算法[J]. 计算机工程, 2022, 48(10):88-94.
- [2] 菊花. 基于改进磷虾群算法的多目标文本聚类方法[J]. 计算机工程与设计, 2022, 43(6):1694-1703.
- [3] 伍建军,马正鹏,杨耀. 基于 J2EE 的项目管理信息系统设计与开发[J]. 制造业自动化, 2022, 44(4):9-12.
- [4] 顾海艳,蒋铜,马卓,等. k-匿名改进算法及其在隐私保护中的应用研究[J]. 信息安全, 2022, 22(10):52-58.
- [5] 李媛媛. 基于区块链技术的财政预算管理效果评估[J]. 自动化技术与应用, 2022, 41(8):45-48.
- [6] FITZPATRICK E K R, B. K L, D. M M, et al. Hospital performance on hospital consumer assessment of healthcare providers and system ratings: associations with nursing factors[J]. Medical Care, 2024, 62(5):288-295.
- [7] 王小鹏,魏统艺,房超,等. 自适应非局部空间约束与 K-L 信息的模糊 C-均值噪声图像分割算法[J]. 控制理论与应用, 2022, 39(7):1261-1271.
- [8] 牟杨. 基于遗传算法的医务人员绩效评价系统[J]. 自动化技术与应用, 2023, 42(1):175-178.
- [9] ABUBEKER K M, BASKAR S. A hand hygiene tracking system with LoRaWAN network for the abolition of hospital-acquired infections[J]. IEEE Sensors Journal. 2023, 23(7):7608-15.

(上接第 162 页)

- [12] 周晶,周丽丽. 基于非侵入式多模态 BCI 的情绪识别与情感分析[J]. 自动化技术与应用, 2024, 43(4):186-188.
- [13] PEREZ-VALERO E, LOPEZ-GORDO M A, VAQUERO-BLASCO M A. An attention-driven videogame based on steady-state motion visual evoked potentials[J]. Expert Systems, 2021, 38(4):12682.
- [14] CHEN Y, YANG C, YE X, et al. Implementing a calibration-free SSVEP-based BCI system with 160 targets[J]. Journal of Neural Engineering, 2021, 18(4):046094.
- [15] 胡晓工,苏志坚,贾晓亮,等. 基于分布式容积信息卡尔曼滤波的锂电池 SOC 估计[J]. 自动化技术与应用, 2024, 43(12):1-7.
- [16] BARACHANT A, BONNET S, CONGEDOM, et al. Multiclass brain-computer interface classification by Riemannian geometry[J]. IEEE Transactions on Biomedical Engineering, 2011, 59(4):920-928.
- [17] KALUNGA E K, CHEVALLIER S, BARTHELEMY Q, et al. Online SSVEP-based BCI using riemannian geometry[J]. Neurocomputing, 2016(191):55-68.
- [18] CONGEDO M, BARACHANT A, BHATIA R. Riemannian geometry for EEG-based brain-computer interfaces; a primer and a review[J]. Brain-Computer Interfaces, 2017, 4(3):155-174.
- [19] LAWHERN J V, SOLON J A, WAYTOWICH R N, et al. EEGNet: A Compact Convolutional Neural Network for EEG-Based Brain-Computer Interfaces[J]. Journal of Neural Engineering, 2018, 15(5):056013.

[10] 刘洪基. 基于混沌 PSO 的大数据智能加权 K 均值聚类算法[J]. 计算机应用与软件, 2022, 39(4):311-319.

[11] NASH A K, WEERAHANDI H, YU H, et al. Measuring equity in readmission as a distinct assessment of hospital performance[J]. JAMA, 2024, 331(2):111-123.

[12] T J B, M C S, REIPING H, et al. ASO visual abstract: association between racial and socioeconomic disparities and hospital performance in treatment and outcomes for patients with colon cancer[J]. Annals of surgical oncology, 2024, 31(2):1095-1096.

[13] 谢忠坚,郭佳欣,刁胤洲,等. 基于机器视觉的兰花几何参数测量方法[J]. 农机化研究, 2025, 47(1):21-25.

[14] 董镇林,伍世虔,叶健,等. 基于自组织聚类和 JS 散度的 RBF 神经网络[J]. 计算机工程与设计, 2024, 45(4):1062-1068.

[15] 李春锋,方春华,侯轩达,等. 基于 K-means 高频局放 10 kV 避雷器快速带电检测方法[J]. 电测与仪表, 2024, 61(7):191-196.

[16] 余建国,肖芳海. 面向电机企业的 CPS 系统设计与开发[J]. 制造业自动化, 2022, 44(6):51-54.

[17] 毛天宇,王星宇,常瑞,等. 面向 Java 语言生态的软件供应链安全分析技术[J]. 软件学报, 2023, 34(6):2628-2640.

作者简介:徐琳(1986—),女,硕士,主治医师,研究方向:卫生管理。

[20] WEIFENG M, CHUANLAI W, XIAOYONG S, et al. MBGA-Net: A Multi-Branch Graph Adaptive Network for Individualized Motor Imagery EEG Classification[J]. Computer Methods and Programs in Biomedicine, 2023(240):107641.

作者简介:张建平(1979—),男,硕士,副研究员,研究方向:自动控制、脑机接口。

通信作者:周丽丽(1980—),女,硕士,研究员,研究方向:类脑人工智能、脑机接口。