

污水处理与医疗大数据交互系统集成研究

曾泳丹¹, 王梓名²

¹四川大学锦江学院 四川成都

²四川大学华西第二医院 四川成都

【摘要】 医疗废水的主要特点是传染性、放射性、毒性和耐药性。近年来, 医疗机构总数、床位数和医生数量均迅速增加, 污水排放量急剧增加。当前阶段, 各大医疗机构针对污水处理, 在相关设施方面做了不断加强, 特别是在污水污染控制方面发挥了一定的作用, 但仍然存在处理设施拥有率低、相应处理水平低、管理不善、不完善等问题。本文充分考虑生态和环境因素安全本文以医疗废水为研究对象, 以大数据信息为研究背景, 利用计算机软件技术和环境科学技术完成了交互式系统的开发。采用 K-means 算法研究了医疗废水处理大数据交互系统的集成问题。分析医疗废水处理的超标程度, 进行二次处理, 提高废水处理率, 通过实验仿真验证算法的有效性, 完成交互系统的集成研究。

【关键词】 废水处理; 医疗处理; 环境; 大数据; K-均值

Research On The Integration Of The Interaction System Of Wastewater Treatment And Medical Big Data

Yongdan Zeng¹, Ziming Wang²

¹Sichuan University Jinjiang College Chengdu, Sichuan

²West China Second Hospital of Sichuan University Chengdu, Sichuan

【Abstract】 Medical wastewater is characterized by infectivity, radioactivity, toxicity and drug resistance. In recent years, the total number of medical institutions, the number of beds, and the number of medical practitioners has increased rapidly, and the amount of wastewater discharged has increased dramatically. At present, the medical institution wastewater facilities have played a positive role in sewage pollution control, but there are many problems such as low ownership rate of treatment facilities, low treatment level, poor management and not fully considering ecological and environmental safety. This paper takes medical wastewater as the research object, takes big data information as the research background, and uses computer software technology and environmental science technology to complete the development of the interactive system. K-means algorithm is adopted to study the integration of big data interactive system of medical wastewater treatment. Analyze the exceeding degree of medical wastewater treatment to cluster, conduct secondary wastewater treatment, improve the treatment rate of wastewater, and verify the effectiveness of the algorithm through experimental simulation, thus completing the integration study of interactive system.

【Keywords】 Wastewater Treatment, Medical Treatment, The Environment, Big Data, K-means

1 前言

医疗废水处理水平一直是衡量医院处理环境质量的重要组成部分。严格的医疗废水处理可以改善医院的诊疗环境, 从而提高医院的治疗效果^[1]。现代医院需要专业的、高质量的医疗废物临时储存室和转运设备, 以及医疗废水净化和处理的空间和设

备, 以确保医院的废物管理和运输以及废水的净化和转运能够准确、有序、质量和数量良好^[2]。为了提高全医院的健康诊疗水平, 建设安全、高效、方便的医疗废水处理中心是现代医院的重要组成部分, 也是医院诊疗环境建设的重要组成部分^[3]。

1.1 污水处理的相关研究

医疗废水主要来源于以下方面: 医疗器械清洗废水、医疗工作(特别是手术)产生的各种传染性

细菌的废水、放射科产生的含放射性元素的废水、医务人员生活废水、患者产生的废水。其中，对门诊、病房、医疗技术、洗衣房和停尸房污水的污染较为严重。在病房的内科、外科、妇科、儿科、皮肤和牙科中，都会有细菌和病毒被排出的废水^[4]。

2 医疗废水处理工艺

2.1 初级加工

医院污水处理的主要污水处理中心通常包括化粪池、沉淀池、双层沉淀池或调节池。在一些小型医院污水处理中心，有些没有沉淀池和调节池，但使用化粪池作为初级处理设施，化粪池废水通过格

栅、计量池（或收集井）直接进入消毒接触池^[5]。一次加工自放电和提升工艺流程如图 1 所示。

2.2 二次过程

当医院污水废水的一级处理不能满足排放标准的水质要求时，采用二级处理工艺。即采用生物处理法去除医院污水中的 COD、苯酚、LAS 等有机污染物，第二个人工处理污水还可通过改善水质，减少消毒剂 的用量，提高消毒效果。医院医用水的二次处理方法有生物接触氧化法、生物旋转法、塔式生物过滤法、氧化沟法^[6]等。典型的过程如图 2 所示。

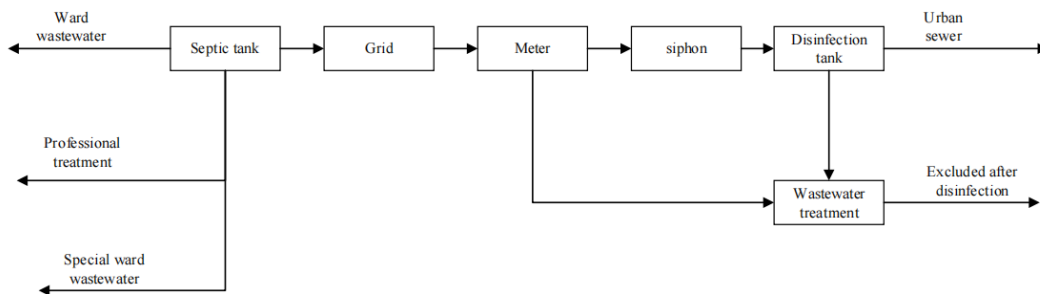


FIGURE 1 Primary wastewater treatment process

图 1 一次加工自放电和提升工艺流程

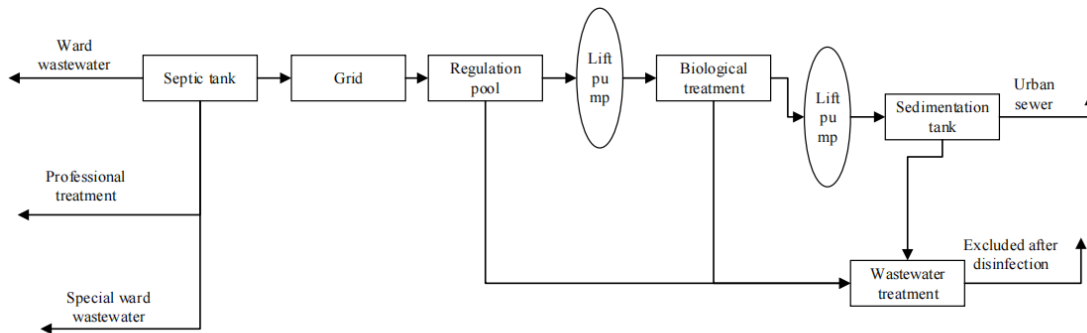


FIGURE 2 Secondary wastewater treatment process

图 2 二级处理工艺

3 医疗大数据研究和系统设计

3.1 基于大数据的相关研究

大数据是指传统数据处理软件在可负担的时间内无法捕获、管理和处理的数据的收集。它的特点是大规模的数据挖掘，但必须依靠云计算、分布式处理、分布式数据库、云存储和虚拟化技术。大数据特点：

(1) 体积，即数据量较大，数据达到 PB 级；

(2) 多样性是多种数据类型，不仅包括传统格式化的数据，还包括未格式化的数据，如来自互联网的视频、图片、日志和地理信息；

(3) 速度，即数据生成和处理速度快；

(4) 准确性是指对数据的定性要求，它会直接影响数据分析的结果；

3.2 医疗废水处理大数据交互系统的集成性研究

在这方面的医疗废水处理研究主要对医疗废水数据进行分析和处理, 这些数据均由底层监测设备收集和传输, 形成废水大数据。废水大数据具有以下四个典型特征:

(1) 为满足综合管理的需要, 所有医院都将在物理空间中连接到系统中。废水监测点数量远远超过工厂数量, 每个监测点将及时实时传输大量废水处理数据;

(2) 多样性、不同的来源、不同的格式;

(3) 医疗废水数据的生成速度非常快;

(4) 大医院排放的废水元素类型可能发生动态变化, 其内容在不同时间不断变化。

3.3 医疗废水大数据交互式系统的设计

本文交互式系统是基于大数据的k-均值算法的设计。k-means 算法是一种聚类算法, 所谓聚类是指根据相似度原则, 将相似度高的数据对象划分为同一聚类, 将相似度高的数据对象划分为不同的聚类^[7]。聚类和分类之间最大的区别是, 聚类过程是一个无监督的过程, 即要处理的数据对象没有任何先验知识, 而分类过程是一个有监督的过程, 即具有先验知识的训练数据集^[8]。通过 k-means 算法, 可以找到由用户指定的数字(k)的簇, 这些簇用质心表示。它具有收敛速度快、原理易于理解、适用于多种数据类型等优点^[9]。该流程图如图 3 所示。

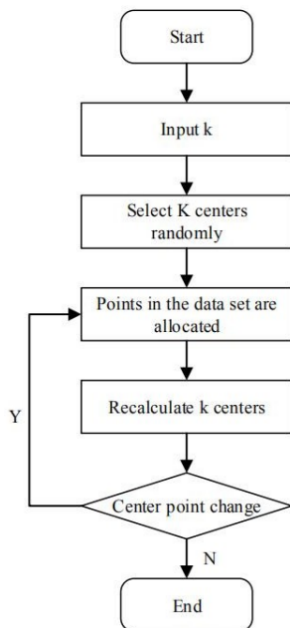


FIGURE 3
Flow chart of k-means

(1) k 值的确定, k 值需要预先确定, 并作为

一个参数输入。

(2) 初始质心选择, 质心的选择是 k-均值中最关键的步骤。选择方式不同, 聚类结果也往往不同。大多数的应用程序都采用了随机选择的方法。由于质心与聚类密切相关, 因此其随机性质心的选择也会导致聚类质量的随机性, 这使得 k-means 算法的聚类效果不稳定。一般来说, 质心是完全随机选择的, 而簇的质量表现又非常低。

(3) 计算成本高。K-means 采用迭代的方法不断更新质心。每次迭代不仅要计算聚类准则函数(目标函数)的值, 还要重新计算下一轮的质心。经过分析, 发现这两种计算的本质是基于两个数据点之间的距离。由于每个污水数据点都是 40 维数据, 被认为是中维, 因此这些数据点之间的距离的计算成本相对较大。因此在废水数据聚类中, k-均值的时间成本较大, 所以需要进一步优化。

3.4 基于大数据的 k-均值算法的优化建模设计

将 k-means 算法应用于医疗废水过度聚类过程中存在的一些问题。根据废水数据的点, 对这些问题进行了适当的改进。首先, 通过预处理去除异常值。其次, 对于维数较高的废水数据, 提出了一种降维、简化计算的方法, 以减少时间开销大的 k-means 算法的缺点^[10]。最后, 在人工选择部分质心的基础上, 结合 k 值, 改进了 k-means 算法, 提出了一种针对废水数据的质心选择方法。采用此人工初始化质心法, 如果指定的 k 等于 2, 则手动选择初始质心, 即一个初始数据中心为最大值表示的数据点, 另一个初始质心为最小值表示的数据点; 如果 k≥2, 可以用在 k=2 选择的方法确定两个初始质心点, 其余 k-2 质心由均值法指定。它能很好地适应 k 值的变化, 并在一定程度上改善了由完全随机初始化质心引起的低质量聚类。

对原污水指标集 A 进行数据预处理, 将 A 中的所有指标与国家标准 S 进行比较, 并进行预处理后进行下一次分析。数据点的处理 B_i 如下, 其中 w 为原始数据的维数, N 为原始数据的条目数:

$$B_i = \left(\frac{A_1 - S_1}{S_1}, \frac{A_2 - S_2}{S_2}, \dots, \frac{A_w - S_w}{S_w} \right)^T = (x_1, x_2, \dots, x_w)^T, (i = 1, 2, \dots, N) \quad (1)$$

提取不超过标准的数据: 经过数据预处理后, 将所有指标都不超过标准的数据进行分组 C0 对不超过标准的医疗废水处理, 其余的用于超标的聚类

分析。其中, 是对一定数据的判断 B_i 如下:

$$\forall_j (B_i^j \leq 0) \in C_0, (j=1, 2, \dots, w, 1 \leq i \leq N) \quad (2)$$

对不合格数据进行聚类预处理: 为每个数据分

$$E = (x_1, x_2, \dots, x_w, ID)^T \xrightarrow{map} F = (x_1, x_2, \dots, x_v, ID)^T, (0 \leq v \leq w)$$

不符合标准的数据集群: 用于映射的数据 F, 在指定 k 后, 采用改进的 k-means 算法进行聚类。

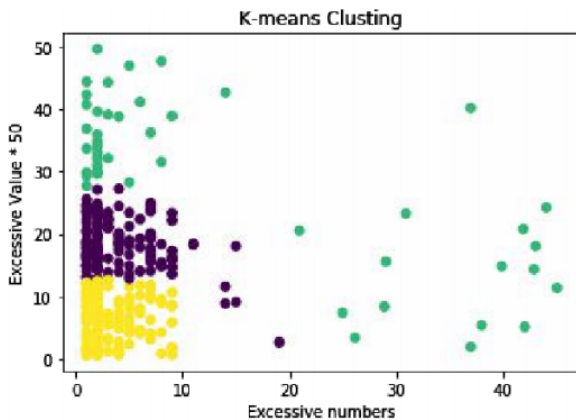


FIGURE 4
Result of simulation

本研究采用一家综合医院获取 2016.08-2018.08 两年实际运行中的废水数据。假设指定的 k 个数为 3, 即对废水的过度标准化程度需要分为轻度超过、适度超过和严重超过。经计算, 最小污水数据点值为 0.01, 最大污水数据点值为 0.73, k-means 算法的结果如图 4 所示。x 坐标为过度标准化号, 范围为 1~47, y 坐标为过度标准化的值 V_x 。从图中可以看出, 顶部和最右边的数据点属于同一个集群, 分别代表的情况的数量超过了超标准和超标准的数量过高, 超级标准的数量太高也属于这个集群, 集群代表一个沉重的过剩。此外, 上述两个数据点之间的紫色数据点簇表示中度过剩。根据超调警告, 可以对这部分水质进行二次废水处理。

4 总结

现代医院医疗废水的管理将越来越受到各部门的重视。现代医院污水处理中心将从单一分散型发展为集中信息型。本文的研究主要以多学科交叉融合为基础, 将生态环境、大数据等学科相结合, 将医疗废水与大数据相结合, 打造综合智能的医疗污水处理系统, 实现了大数据在废水领域的应用, 为污水处理研究奠定了理论基础。

配一个唯一的标签 ID, 并将所有标记的聚类数据 E 映射到 F。式中, v 为映射后的新数据维, 即 0~w 之间值不确定的不合格污水指数:

参考文献

- [1] Sivarajah, U., Kamal, M.M., Irani, Z., Weerakkody, V. (2017) Critical analysis of Big Data challenges and analytical methods. *Journal of Business Research*. 70, 263-286.
- [2] Akter, S., Wamba, S.F. (2017) Big data and disaster management: a systematic review and agenda for future research. *Annals of Operations Research*. 9, 1-21.
- [3] Athey, S. (2017) Beyond prediction: using big data for policy problems. *Science*. 355, 483-485.
- [4] Xu, W.C., Zhou, H.B., Cheng, N., Lyu, F., Shi, W.S., Chen, J.Y., Shen, X.M. (2018) Internet of Vehicles in Big Data Era. *IEEE/CAA Journal of Automatica Sinica*. 5, 19-35.
- [5] Bajari, P., Chernozhukov, V., Hortacsu, A., Hortacsu, A., Suzuki, J. (2018) The Impact of Big Data on Firm Performance: An Empirical Investigation. *Nber Working Papers*, National Bureau of Economic Research, Inc.
- [6] Yang, W.X., Li, L.G. (2018) Efficiency evaluation of industrial waste gas control in China: A study based on data envelopment analysis (DEA) model. *Journal of Cleaner Production*. 179, 1-11.
- [7] Song, Z.J., Liu, H.M., Meng, F.X., Yuan, X.Y., Feng, Q., Zhou, D.W., Romaní, J.R.V., Yan, H.B. (2019) Zircon U-Pb Ages and Hf Isotopes of Neoproterozoic Meta-Igneous Rocks in the Liansandao Area, Northern Sulu Orogen, Eastern China, and the Tectonic Implications. *Journal of Earth Science*. 30(6), 1230-1242.
- [8] Choi, E.S., Min, K., Kim, G.J., Kwon, I., Kim, Y.H. (2017) Expression and characterization of Pantoea CO dehydrogenase to utilize CO-containing industrial waste gas for expanding the versatility of CO dehydrogenase. *Sci Rep*. 7, 44323-44324.

- [9] Guo, S.K., Chen, R., Li, H., Zhang, T.L., Liu, Y.Q. (2019) Identify Severity Bug Report with Distribution Imbalance by CR-SMOTE and ELM. International Journal of Software Engineering and Knowledge Engineering. 29, 139-175.
- [10] https://blog.csdn.net/weixin_42137700/article/details/90230443.

收稿日期: 2022 年 9 月 18 日

出刊日期: 2022 年 10 月 25 日

引用本文: 曾泳丹, 王梓名, 污水处理与医疗大数据交互系统集成研究[J]. 国际计算机科学进展, 2022, 2(3): 34-38.

DOI: 10.12208/j. aics.20220042

检索信息: RCCSE 权威核心学术期刊数据库、中国知网 (CNKI Scholar)、万方数据 (WANFANG DATA)、Google Scholar 等数据库收录期刊

版权声明: ©2022 作者与开放获取期刊研究中心 (OAJRC) 所有。本文章按照知识共享署名许可条款发表。 <http://creativecommons.org/licenses/by/4.0/>



OPEN ACCESS