

基于边缘计算的医疗资源配置重构优化模型构建

夏甫开提·阿力甫¹, 周京涛², 努尔比亚吾素因², 买买提·牙森¹

1. 新疆医科大学公共卫生学院, 新疆 乌鲁木齐 830054; 2. 新疆医科大学第七附属医院, 新疆 乌鲁木齐 830011

【摘要】为了缓解医院环境中医疗资源不足的问题,本文提出一种基于边缘计算的医疗资源配置重构优化模型。通过使用多准则决策分析方法动态优化网络边缘侧的计算资源,以避免将任务下沉到云端中的数据安全,患者因素或者高延迟等问题。最后进行实验评估提出的重构优化和资源分配方法,结果表明本文方法在分配基于边缘的资源执行任务时的有效性。

【关键词】网络边缘; 医疗资源; 云端; 重构优化

【中图分类号】R197; R318

【文献标志码】A

【文章编号】1005-202X(2022)11-1407-05

Reconfiguration and optimization model for medical resource allocation based on edge computing

Xiafukaiti·Alifu¹, ZHOU Jingtao², Nurbiawusuyin², Maimaiti·Yasen¹

1. School of Public Health, Xinjiang Medical University, Urumqi 830054, China; 2. The Seventh Affiliated Hospital of Xinjiang Medical University, Urumqi 830011, China

Abstract: A reconfiguration and optimization model for medical resource allocation based on edge computing is proposed for alleviating the problem of inadequate medical resources in the hospital. The multi-criteria decision analysis method is used to dynamically optimize the computing resources at network edge, thereby avoiding the problems that the tasks sink into the cloud data security and patient factors or eliminating the problem of high delays. Finally, the proposed reconfiguration optimization and resource allocation method is evaluated through experiments, and the results show that the proposed method is effective in allocating resources based on edge computing when carrying out tasks.

Keywords: network edge; medical resource; cloud; reconfiguration and optimization

前言

近年来,物联网和人工智能的持续发展为医疗保健行业带来了新的机遇和长足的发展^[1-6]。传感器和驱动等硬件技术的进步促进了医疗与物联网的结合,即医疗互联网,这种新兴的范式为医疗设备、应用程序和医疗保健服务的完美闭环连接提供了最基础的设施^[7-10]。医疗物联网旨在通过集成医疗保健数据、流程和医疗设备或移动端的医疗保健应用来不断改善医疗保健服务。随着能够捕获和传输医疗数据的医疗设备和医疗系统数量的迅速增加,医疗物联网也相应地快速发展。因此,在未来几年医疗

物联网很可能在提高医疗保健机构的运营生产率,以及医疗诊断和治疗的速度和准确性等方面发挥越来越重要的作用^[11-15]。然而,一般的医疗设备在处理高级人工智能或机器学习相关任务时的计算能力相当有限,这也是要建立有效的健康信息工作流程方面的巨大挑战。由此看来,医疗设备可能需要将某些已经超越其本身硬件能力的高级智能任务转移到更强大的计算资源,以保持医疗诊断或治疗以及患者实时监测所需的速度和准确性^[16]。此时为了完成任务,可以对任务进行下沉操作。然而,通常在下沉医疗物联网任务时会导致相应医疗保健服务的性能下降,经过分析,这或许与涉及医疗数据(如患者和相关性能数据)的网络安全或漏洞有关^[17]。此外,医疗物联网任务在网络传输时可能会受到延迟问题的影响,这可能会导致完成任务的响应时间更长。因此,需要首先确定运行在医院的医疗物联网系统上的任务的可下沉性。具体来讲,任务的可下沉性被定义为物联网系统需要确定一个任务是否具有可以被下沉到类似或其它不同网络中的一个或多个资源

【收稿日期】2022-05-13

【基金项目】新疆维吾尔自治区自然科学基金(2021D01A175)

【作者简介】夏甫开提·阿力甫,硕士,研究方向:卫生资源配置, E-mail: alipu1996@163.com

【通信作者】买买提·牙森,硕士,教授,博士生导师,研究方向:医学教育、卫生资源, E-mail: nur4056@163.com

的能力。对于这种情况,医院内的医疗物联网系统确定哪些任务可以下沉,哪些任务需要在本地环境中处理才可以。然而,确定医疗任务可下沉性的问题往往涉及多个相互冲突的准则,即该过程通常涉及到不同约束的许多决策变量。因此,需要将医疗物联网系统中的任务可下沉性视为一个多准则的决策过程。

医院中的一些医疗保健应用程序一般通过局域网运行^[18]。某些特殊的应用程序或者高级手术可能需要在本地网络范围之外处理某些任务(如云端)。由此便引入了云计算的概念。云计算将巨大的数据计算处理程序分解成无数个小程序,然后,通过多部服务器组成的系统进行处理和分析这些小程序得到结果并返回给用户。云计算的核心概念就是以互联网为中心,在网站上提供快速且安全的云计算服务与数据存储,让每一个使用互联网的人都可以使用网络上的庞大计算资源与数据中心。云计算的这种特性涉及到两个问题:其一,医疗数据必然伴有隐私性,而这种将数据上传至云端的方式必然会带来一些潜在的数据外泄和网络安全问题;此外,一些手术在高延迟的情况下必然会有风险,甚至会威胁到患者的生命。这显然带来了医疗资源如何配置的问题。此问题通常可以使用边缘计算来解决,它将网络、存储和应用处理能力等计算资源转移到更接近用户设备的边缘侧,从而就近提供智能服务。边缘计算的高安全和低时延特性也使其成为在医院环境中解决某些特殊高级任务的理想解决方案。然而,与云计算相反,边缘计算环境的资源如计算、存储和网络等通常相当有限^[6]。

基于上述原因,本研究构建一种基于边缘计算的医疗资源配置重构优化模型(ROME)以有效地提供计算资源,进而更有效地支持医疗物联网,使其在医院可以更有效地对计算资源进行分配。具体来说,为了缓解医院环境中边缘计算资源的不足,本研究引入了一种优化策略,将资源分配视为一个多准则决策问题。通过优化完成计算任务所需的资源,可以更有效地将数据的处理分配到不同的医院环境中,并确定将任务转移到外部资源(如云端)的可下沉性。ROME 可以集成到医院本部的 IT 基础设施中,而由于在网络边缘侧运行医疗保健相关操作的需求日益增加,因此为了能更有效地维护医疗保健应用程序和服务以满足需求,同时有效满足响应时间,也可以将其放置于附近的微型数据中心。

1 方法

网络边缘侧的医疗物联网相关设备的激增需要

更有效的技术来分散数据处理任务或涉及患者数据的工作负载。为了避免将患者或性能数据发送到云资源的弊端,需要在医院中配备执行高级计算操作所需的边缘节点,进而在边缘端处理任务,图 1 为本研究的 ROME 框架。

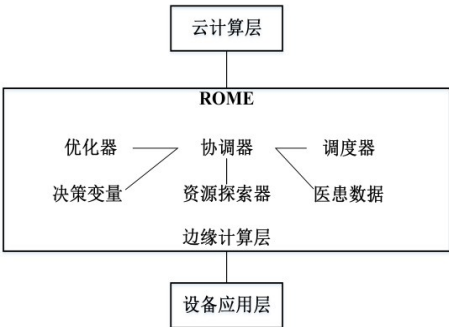


图 1 整体框架图
Figure 1 Overall flowchart

如图 1 所示,最底下是应用设备层,比如手术机器人、护理设备、临床可穿戴设备和医疗传感器等。中间是 ROME 框架,它位于边缘计算层,可以位于医院内部或靠近医院环境的微型数据中心。在边缘计算层中处理相关工作任务有许多优点,最显著的是可以减少医疗诊断和治疗所需的时间,以及可以更快访问患者的数据,提高对患者和性能数据的可访问性(这些数据可用于进行保险理赔等)。

由于医疗物联网包含了许多技术,因此在进行工作时可能需要应用程序处理通过医疗物联网设备收集的大量实时数据。为了加快医疗物联网收集数据的过程,优化医院中的资源分配势在必行,需要边缘环境之外额外资源的任务或操作可以被下沉到云端进行进一步处理。这样一来,有必要确定哪些任务可以在本地环境中运行,哪些任务需要上移到云端。

ROME 由许多组件组成,这些组件通过在任何下沉决策之前以分散的方式优化基于边缘的资源来改善确定任务可下沉性的流程。ROME 框架包含一个协调器,用于控制和管理边缘计算层中的现有资源,它会创建一个矩阵,优化器使用该矩阵确定医院环境中的最佳资源分配计划。考虑到下沉请求经常涉及多个相互冲突的准则,ROME 采用了多准则决策分析方法中的逼近理想解排序法(TOPSIS)。其中的决策变量用于设置优化流程,并使管理员能够识别每个决策变量的重要程度。ROME 预定义一些工作类型,以服务于多种类型的医疗物联网应用程序。具体来说,本研究确定了 4 种主要的医疗物联网工作负载类型:CPU 优先、内存优先、计算优先和成本优

先。这样一来,例如延迟敏感的应用程序可以与相应的工作类型相关联,以确保该类别中的任务比其它非关键任务具有更高的优先级。资源探索用于寻找边缘计算层中的现有资源,它由许多边缘计算节点组成。网络边缘的资源根据硬件能力的不同而产生差异变化,包括处理单元、内存、磁盘读/写速率、网络接收/发送速率等。调度器用于识别基于云端的计算资源,将可下沉的任务分配到其中。

2 评估和结果

为了评估 ROME 的性能,在 gwa-t-12 bitbrains 数据集上进行了实验^[19]。该数据集包含来自分布式资料中心的 1 250 个虚拟机的相关属性,同时提供了与映射边缘计算资源相关的跟踪数据,如处理器提供的容量,处理器使用率(单位:MHz),内存容量(单位:kb),内存使用率(单位:kb),磁盘读/写速率(单位:kb/s)和网络发送/接收速率(单位:kb/s)。此外,本研究还使用这些属性另外派生出两个可以代表模型的成本和能耗属性。对成本属性,使用 Intel Xeon-8260 处理器(2.4 GHz, 165 W, 24 核),使用文献[20]中的 IT 基础设施功率计算器,并将配置项映射到 gwa-t-12 bitbrains 的属性中。这样一来,便可计算出系统的总功率,然后利用这些信息推导出一个成本模型,并根据文献[21]导出与所需的边缘计算资源消耗相关的能耗属性。

通过使用派生出的成本属性和能耗属性,再进行多准则决策分析^[22]。具体来讲,在定义决策变量的基础上,采用 TOPSIS 进行决策,从而识别出可用资源,以完成边缘侧的工作。在边缘计算环境中资源不足的情况下,ROME 可以将任务下沉到医院附近的另一个边缘层(如附近的微型数据中心)或云端等其他资源中。

在设计任务的计算下沉策略时,本研究主要关注边缘计算资源的高效与动态分配。也就是说,通过在网络边缘中有效地分配资源,就有可能优化医疗物联网中应用程序的性能,从而增强下沉过程。由于任务可以根据需要分配的硬件资源的需求而变化,因此确定了代表任务或操作类型的若干类别。表 1 给出医院环境中可能存在的各种基于边缘的医疗物联网操作类型。

不同的任务类型可表示与特定属性相关的优先级,如图 1 所示。假设应用程序可以通过 ROME 的协调器组件,在请求分配基于边缘计算的资源时已经提供任务类型,然后使用一些实例评估 Edgify 医疗资源配置重构优化模型。

表 1 基于边缘计算的医疗任务类型

Table 1 Types of medical tasks based on edge computing

任务类型	具体描述
计算优先	需要大量数据处理的关键且时间敏感型任务,如需要高级人工智能或深度学习算法处理的癌症诊断工作
成本优先	非关键但时间敏感型任务,需要在 CPU、内存以及最重要的成本方面进行高效处理,如为特定病例确定可能的治疗方案
CPU 优先	类似于计算优先任务,除了成本不是决策时的主要因素,并需要足够的内存执行任务
内存优先	涉及处理通过网络媒体传入数据的任务或处理涉及非永久性存储器的高级数据结构的任务
一般任务	不需要在 CPU、内存、磁盘 I/O、连通性、成本或能耗等方面有特定需求的任务,所有的决策变量都被一视同仁

2.1 计算优先和成本优先的实例

计算优先的操作通常是关键且时间敏感的任务。在医疗物联网的大架构下,关键任务包括使用对医院/患者至关重要的系统或设备。这种任务对操作的持续性要求是必不可少的,任何扰乱或中断都可能对医院的经营或对正在进行手术的病人造成严重损害。计算优先操作的典型例子,如一个由手术机器人进行的手术,它需要捕捉图像进行进一步处理,并向由医生组成的手术团队提供建议,以便在手术室中对患者做出相当精确的判决与操作。因此在执行计算优先的任务时,成本和能耗并不是优先考虑的因素。

成本优先的任务有点类似于计算优先任务,但与之不同的是此时成本和能耗会在决策中扮演着更重要的角色。重要性或优先级将作为 TOPSIS 决策方法的权重。本研究利用 gwa-t-12 bitbrains 数据集进行了相关实验。TOPSIS 生成了一个优化分数(或性能分数),称之为 ROME 分数^[23]。假设 1 250 个虚拟机分布在基于边缘计算的医院环境中,并且此时 ROME 遇到一个请求分配资源的任务。ROME 使用 TOPSIS 动态地识别合适的资源,并对执行任务时可以使用的最佳或最理想的资源做出决策。基于此,对计算优先和成本优先的任务类型进行了请求资源的实验。使用相同的边缘计算节点比较总体性能分数,即 ROME 分数。为了便于比较,实验中还包含了一般情况下的任务类型。ROME 可以识别出可以满足不同任务需求的可用资源。也就是说,在执行医疗任务时可使用 ROME 分配最佳的可用计算资源。可以看到,对于计算优先的任务,ROME 识别出边缘计算节点 e6 是执行该任务最合适的资源。类似地,它也识别出了成本优先的任务也最适合 e6。表 2 显示在 gwa-t-12 bitbrains 数据集中与边缘节点相关的一些归一化数值。

表2 计算优先任务的边缘节点资源容量/承载力

Table 2 Resource capacity/carrying capacity of edge nodes for calculation priority tasks

边缘计算节点	CPU 使用率	可用内存	成本	能耗	ROME 分数
e1	0.48	0.63	0.50	0.41	0.87
e2	0.18	0.18	0.07	0.16	0.69
e3	0.15	0.34	0.09	0.23	0.92
e4	0.37	0.34	0.32	0.35	0.68
e5	0.07	0.05	0.03	0.20	0.68
e6	0.13	0.38	0.07	0.22	1.00
e7	0.50	0.20	0.54	0.43	0.34
e8	0.23	0.29	0.17	0.28	0.77
e9	0.29	0.19	0.23	0.31	0.55
e10	0.42	0.17	0.39	0.37	0.37

对于计算优先的任务,对CPU使用率和可用内存分别采用45%的让行率,而对成本和功耗分别采用10%的让行率。从表2可以看出,边缘计算节点e6的ROME分数最高,其次是边缘计算节点e3。边缘计算节点e6属于CPU利用率和可用内存都较低边缘计算资源。因此,本研究构建的资源配置重构优化策略可以推荐一个可以平衡CPU和内存的解决方案。此外,边缘计算节点e1在所有ROME的分数中排名第3,因为它有最多的可用内存,同时它也有52%的CPU可用率。边缘计算节点e1在ROME框架用于计算优先的任务时排名第3,但在成本优先的任务中排名第9,因为它是第2昂贵的可用资源(成本属性为0.50,仅稍好于排名最后的e7)。

2.2 CPU 优先、内存优先、成本优先以及能耗优先的任务

使用其它的一些主要关注某个中心因素或属性实例评估ROME。如CPU优先的任务类型关注CPU使用率,将其作为与其他属性关联的权重最小或没有权重的主要属性。在这个实验中,比较了成本优先任务、CPU优先任务、内存优先任务、能耗优先任务以及一般任务的结果,实验结果见图2。

同样地,也考虑了与上个实验中相同的边缘节点集,以说明ROME优化模型的有效性。如图2所示,与表2中提供的数据相一致,边缘计算节点e1具有最多的内存优先的分段。与其他的边缘节点相比,它具有0.63的最高可用内存。若考虑到CPU优先的任务,ROME将会标识到边缘计算节点e5作为最优计算资源。此结果也与表2中的数据一致,即在所有的边缘计算资源中,边缘计算节点e5的CPU使用率最低,仅为7%。此外,ROME根据ROME分数

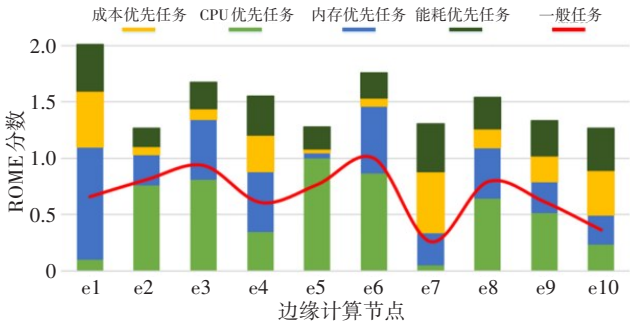


图2 基于成本优先任务、CPU优先任务、内存优先任务、能耗优先任务以及一般任务下的边缘计算节点资源分配结果

Figure 2 Resource allocation results of edge computing nodes in cost priority tasks, CPU priority tasks, memory priority tasks, energy priority tasks and general tasks

在计算资源排名列表中识别e6、e3、e2。这些结果与表2中的数据一致,其中e6、e3和e2的CPU使用率分别为0.13、0.15和0.18。与内存优先的任务不同,对于CPU优先的情况,ROME尝试寻找具有较低CPU使用率的资源。

3 结论

本研究构建了基于边缘计算的医疗资源配置重构优化计算模型ROME。具体来说,在进行涉及患者数据的医疗保健应用程序的任务下沉时,由于该过程经常涉及如隐私、延迟、成本、网络可靠性等多个相互冲突的准则,因此将其设计为一个多准则决策过程。通过一系列实验评估ROME,结果证明本研究方法对于医疗计算资源配置重构优化的有效性。

【参考文献】

[1] 张杰,张鞠成,徐洪良,等. 物联网在医疗设备管理领域的应用与发展[J]. 医疗装备, 2022, 35(7): 186-189.
Zhang J, Zhang JC, Xu HL, et al. Application and development of the internet of things in the field of medical device management [J]. Chinese Journal of Medical Device, 2022, 35(7): 186-189.

[2] 邵彩萍. 5G+智慧医疗服务平台的交互设计与具体产品应用研究[J]. 湖北工程学院学报, 2022, 42(2): 186-189.
Shao CP. Research on interaction design of 5G+smart medical service platform and application of specific products [J]. Journal of Hubei Engineering University, 2022, 42(2): 186-189.

[3] 陈军,于大江,彭长林. 基于物联网的医疗设备综合管理系统设计与实现[J]. 中国医学装备, 2022, 19(4): 134-138.
Chen J, Yu DJ, Peng CL. Design and implementation of an integrated management system for medical equipment based on the internet of things [J]. China Medical Equipment, 2022, 19(4): 134-138.

[4] Sun L, Jiang X, Ren H, et al. Edge-cloud computing and artificial intelligence in internet of medical things: architecture, technology and application [J]. IEEE Access, 2020, 3(9): 1-16.

[5] Tang Q. Engineering design of electronic fence system based on intelligent monitoring and wireless local area network [J]. Alexandria Eng J, 2021(3): 87-95.

[6] Padilla-Cuevas J, Reyes-Ortiz JA, Bravo M. Ontology-based context event representation, reasoning, and enhancing in academic environments [J]. Future Internet, 2021, 13(6): 151-168.

- [7] Mondal MA, Rehena Z. Stacked LSTM for short-term traffic flow prediction using multivariate time series dataset[J]. *Arabian J Sci Eng*, 2022: 1-15.
- [8] Katsumata S, Yamada S, Yamakawa T. Tighter security proofs for GPV-IBE in the quantum random oracle model[J]. *J Cryptol*, 2021, 34(1): 18-26.
- [9] Zhou B, Wu Q, Zhao X, et al. Construction of 5G all-wireless network and information system for cabin hospitals[J]. *J Am Med Inform Assoc*, 2020(6): 19-28.
- [10] Gao Y, Zhang H, Yu F, et al. Joint computation offloading and resource allocation for mobile-edge computing assisted ultra-dense networks[J]. *J Commun Inform Netw*, 2022(1): 7-19.
- [11] 卢松龄. 人工智能与物联网对医疗产业的影响[J]. *黑龙江科学*, 2022, 13(20): 98-100.
Lu SL. The impact of artificial intelligence and internet of things on medical industry[J]. *Heilongjiang Science*, 2022, 13(20): 98-100.
- [12] Wang C, Zhou T. Multi-level storage based auditing scheme for 5G and beyond defined edge computing[J]. *J Surveill Secur Saf*, 2022, 3(1): 16-26.
- [13] Cui X, Envelope R. Application of intelligent edge computing technology for video surveillance in human movement recognition and Taekwondo training[J]. *Alexandria Eng J*, 2022, 61(4): 2899-2908.
- [14] Ju X, Gou R, Xiao Y, et al. The use of edge computing-based internet of things big data in the design of power intelligent management and control platform[J]. *Int J Grid Utility Comput*, 2022(1): 13.
- [15] Song S, Fang Z, Jiang J. fast-DRD: fast decentralized reinforcement distillation for deadline-aware edge computing[J]. *Inform Process Manag*, 2022(2): 59.
- [16] Li Y. Groundwater level change based on edge computing and translation accuracy of urban English (Retraction of Vol 14, art No.1682, 2021)[J]. *Arabian J Geosci*, 2021(22): 14.
- [17] Hu J, Chen C, Cai L, et al. UAV-assisted vehicular edge computing for the 6G internet of vehicles: architecture, intelligence, and challenges[J]. *IEEE Commun Stand Mag*, 2021. DOI: 10.1109/MCOMSTD.001.2000017.
- [18] Hsu CH, Alavi AH, Dong M. Guest editorial: special issue on the "Sustainable Development of Smart Cities with Edge Computing Techniques"[J]. *J Internet Technol*, 2021(5): 22.
- [19] Zhou C, Yu Y, Yang S, et al. Intelligent immunity based security defense system for multi-access edge computing network[J]. *China Commun*, 2021, 18(1): 100-107.
- [20] Demler M. Rockchip accelerates AI for IOT[J]. *Microprocessor Rep*, 2019, 33(7): 27-29.
- [21] Zhao T, Liu Y, Shou G, et al. Joint optimization of latency and energy consumption for mobile edge computing based proximity detection in road networks[J]. *China Commun*, 2022, 19(4): 19-28.
- [22] Zheng S, Yuan H, Xin S. Computation offloading and resource allocation algorithm for collaborative LEO satellite multi-access edge computing[J]. *Acta Electronica Sinica*, 2022, 50(3): 567-573.
- [23] Yang Y, Ding S, Liu Y, et al. Fast wireless sensor for anomaly detection based on data stream in an edge-computing-enabled smart greenhouse[J]. *Digital Commun Networks*, 2022, 8(4): 10.

(编辑:黄开颜)