

信息年龄和能耗联合优化的无线体域网卸载和调度策略

张政^{1,2}, 谢鑫³, 柏桐^{2,4}, 林金朝², 李章勇⁴

(1.重庆邮电大学通信与信息工程学院, 重庆 400065;

2.重庆邮电大学光电信息感测与微系统重庆市重点实验室; 重庆 400065;

3.重庆邮电大学自动化学院, 重庆 400065; 4.重庆邮电大学光电工程学院, 重庆 400065)

摘要: 在无线体域网 (WBAN) 中, 为保障生理数据的实时传输并降低节点能耗, 提出了一种联合优化的卸载与调度策略。该策略通过选择在传感节点或 Sink 处理数据, 最小化信息年龄 (AoI) 与能耗的加权和。为解决卸载和调度决策的强耦合问题, 采用双层马尔可夫决策过程 (MDP), 并通过深度强化学习 (DRL) 应对维数灾难。仿真结果表明, DRL 策略在性能上接近 MDP, 且在节点增加时, 相较于 RRG 与 EG 策略, 平均加权和分别降低约 3.58% 和 24.9%, 收敛速度约为 MDP 的 2 倍。

关键词: 无线体域网; 卸载调度; 信息年龄; 能耗

中图分类号: TN915.04

文献标志码: A

DOI: 10.11959/j.issn.1000-436x.2024148

Joint optimization strategy of age of information and energy consumption for offloading and scheduling in WBAN

ZHANG Zheng^{1,2}, XIE Xin³, BAI Tong^{2,4}, LIN Jinzhao², LI Zhangyong⁴

1. School of Communications and Information Engineering, Chongqing University of Posts and Telecommunications, Chongqing 400065, China

2. Chongqing Key Laboratory of Optoelectronic Information Sensing and Microsystems, Chongqing University of Posts and Telecommunications, Chongqing 400065, China

3. School of Automation, Chongqing University of Posts and Telecommunications, Chongqing 400065, China

4. School of Optoelectronic Engineering, Chongqing University of Posts and Telecommunications, Chongqing 400065, China

Abstract: Both real-time physiological data transmission and reduced energy consumption are critical to wireless body area network (WBAN). A joint optimization strategy for offloading and scheduling was proposed to minimize the weighted sum of age of information (AoI) and energy consumption by determining whether data should be processed at sensor nodes or the Sink. To handle the strong coupling between offloading and scheduling decisions, a two-layer Markov decision process (MDP) was used to approximate the optimal solution. A deep reinforcement learning (DRL) approach was introduced to address the dimensionality issue. Simulations show that the DRL strategy performs comparably to the MDP under various weight factors and frame lengths. Furthermore, as the number of sensor nodes increases, the DRL strategy reduces the weighted sum by 3.58% and 24.9% compared to RRG and EG strategies, respectively, and converges twice as fast as the MDP.

Keywords: wireless body area network, offloading and scheduling, age of information, energy consumption

收稿日期: 2024-02-29; 修回日期: 2024-07-26

通信作者: 李章勇, lizy@cqupt.edu.cn

基金项目: 国家自然科学基金资助项目 (No.U21A20447, No.62171073, No.62311530103); 重庆市自然科学基金资助项目 (No.CSTB2022NSCQ-MSX1523); 重庆邮电大学博士研究生创新人才基金资助项目 (No.BYJS202206)

Foundation Items: The National Natural Science Foundation of China (No.U21A20447, No.62171073, No.62311530103), The Nature Science Foundation of Chongqing (No.CSTB2022NSCQ-MSX1523), The Innovative Talent Program for Doctoral Students of Chongqing University of Posts and Telecommunications (No.BYJS202206)

0 引言

无线体域网 (WBAN, wireless body area network) 作为一种关键的可持续医疗解决方案, 正被越来越广泛地应用^[1]。WBAN系统由一个Sink和 n 个分布在人体周围或内部的无线传感节点组成, 传感节点以生物医学信号的形式记录人体生理数据的变化, 并通过无线信道将其传输到Sink, 以帮助病人进行实时、连续和长时间的监测^[2], 同时Sink可以进一步将这些生理数据传输到远程处理中心, 以便为病人提供及时的诊断和医务人员的建议。WBAN不同于一般的网络, 因为涉及生理数据的传输, 目前WBAN的各种通信指标受到更多的关注, 尤其是生理数据传输的实时性和传感节点的能耗^[3]。在实际情况中, 过期的生理数据不仅可能导致不必要的能耗, 而且还可能造成严重的医疗事故。例如, 过期的生理数据可能会影响医疗决策, 不可预见的紧急情况发生时可能会造成难以想象的后果^[4]。WBAN必须保证生理数据的时效性, 而一般的时延指标已不能满足WBAN特殊的时效性。为了衡量生理信息的新鲜度和及时性, 信息年龄 (AoI, age of information) 被提出作为一个新的度量标准, 因为AoI不仅考虑了更新生理数据的传输时间, 还考虑了队列中的等待时间和接收后的停留时间^[5], AoI的广泛应用有效地解决了生理数据包传输中排队时延对信息及时性的影响^[6]。文献[7]首次提出了AoI的概念, 随后不少文献通过引入AoI来评估无线网络的性能。文献[8]考虑了一种时变信道下最小化平均AoI的调度策略。文献[9]提出了一种传感节点协同调度方法来考虑传感数据的新鲜度。文献[10]通过对时间敏感用户进行混合流量设计并提供了一种低复杂的调度策略用于最小化平均AoI以及满足系统吞吐量的约束。文献[11]考虑了一种传输调度策略最小化信息加权和平均年龄。在不同的数据更新类型下, 文献[12]研究了噪声信道中与AoI相关的链路调度策略。文献[13]研究了物联网中具有最后期限的平均峰值AoI。这些关于AoI的研究对网络系统性能的提升有所帮助, 但鲜有考虑WBAN的场景, 在受限的传感节点能耗的情况下, 如何解决时效性和能耗的最优化问题更是需要解决的重点。

另外, 在实际的WBAN系统中, 现有的Sink调度策略没有考虑生理数据的暂态值, 而是更加关注WBAN系统的吞吐量和能量效率。Kuang等^[14]分析

了本地计算、边缘计算和单源-目标网络中部分卸载的通信指标。Song等^[15]开发了具有确定性信道的多用户网络中的联合部分卸载和调度策略来提高系统性能。Samanta等^[16]提出了一种动态链接建立和协作调度方案来最大限度地降低数据包传输时延并最大化网络吞吐量。Kim等^[17]设计了一种WBAN中基于优先级的自适应调度和功率控制方案来解决WBAN中多目标优化问题 (MOP, multi-objective optimization problem), 同样, 文献[18]通过构建优先级规则的排队模型来调度观测到的生理数据包随机到达每个Sink的问题。文献[19]采用混合调度的方法设计了一种紧急优先时隙分配方案来处理不同情况下的紧急生理数据。文献[20]通过联合优化调度来限制能量和时间速率、计算调度和发射功率设计来分析所考虑网络的性能。这些基于调度的策略能够解决一部分生理数据的处理, 但是更多无用的生理数据传输和Sink处理导致更多的能耗, 并且时延的问题始终存在。在噪声信道下, 如何卸载生理数据和链路调度来保证时效性和能耗的需求还有待研究。

为了解决以上问题, 本文联合设计了一种二进制卸载和调度策略使信息年龄和能耗的平均加权和最小化, 主要贡献总结如下。

1) 引入新鲜度作为WBAN系统衡量指标, 更加全面地考虑生理数据带来的时延。为了最小化AoI和能耗平均加权和, 同时将联合问题最优化为一个双层马尔可夫决策过程 (MDP, Markov decision process), 然后通过迭代方法逼近最优解。

2) 为了克服MDP策略的维数灾难, 本文提出了一种基于深度强化学习 (DRL, deep reinforcement learning) 策略来产生合理的卸载模型和调度决策。同时考虑到卸载人工神经网络 (ANN, artificial neural network) 与调度ANN之间的异质性, 采用后期存储方法使2种神经网络能够同时训练。仿真结果表明, DRL策略的性能接近最优。

3) 由于MDP策略调度和卸载决策的执行时间和依赖系统状态的不同, 本文DRL策略同步训练2个异构决策神经网络, 仿真结果显示出较好的性能。

1 系统模型和优化问题

本文考虑一个由可计算传感节点和Sink组成的WBAN系统模型, 如图1所示。假设系统是基于帧

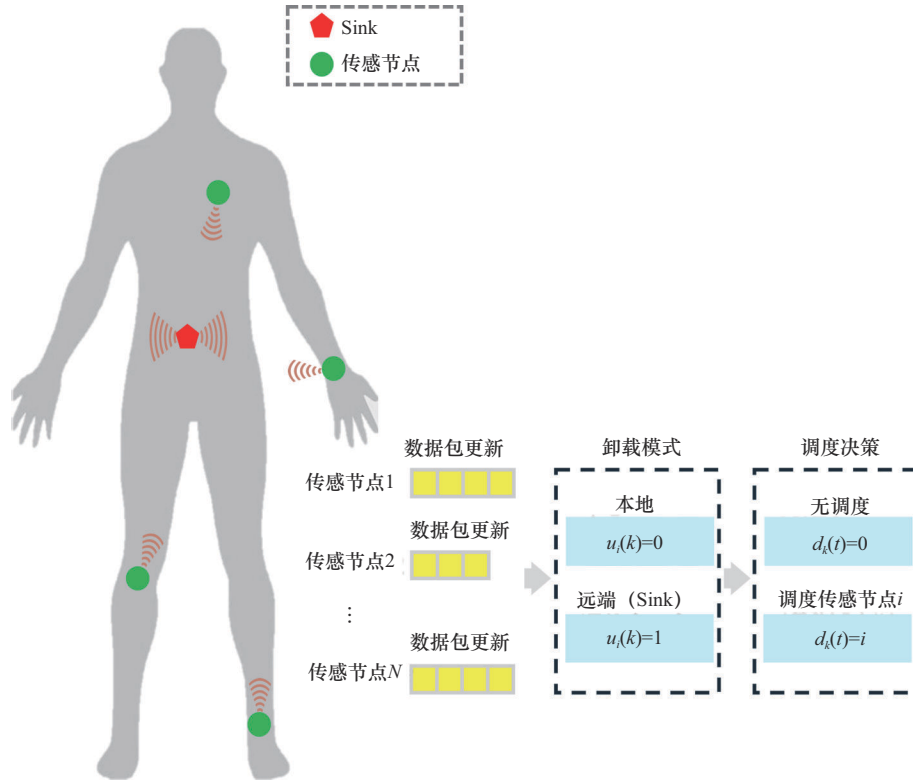


图1 系统模型

的并且每个帧包含 T 个时隙。在每帧的开始，传感节点 $i \in \{1, 2, \dots, M\}$ 采样相关的生理数据并获得 $L_i \geq 2$ 个状态更新的数据包（如脑电信号或心率信号等）。传感节点 i 可以选择本地卸载模式或 Sink 处理模式。其中，本地卸载模式后会将生理数据包传输到 Sink 或直接将生理数据包卸载到 Sink 进行处理。由于本地卸载后的生理数据包更新大小一般认为较小，为了简单起见，本文将之视为一个数据包。假设在每个时隙中最多选择一个传感节点 i 通过无线信道将一个生理数据包传输到 Sink，如果之前帧的状态更新未被传输，则将由新的状态进行更新。

1.1 卸载和调度模型

在每个帧开始生成一个新的状态时，传感节点 i 确定其卸载模式，并且在每个时隙开始时做出调度策略。传感节点 i 进行本地卸载模式时，由于计算能力有限需要多个时隙进行数据处理；Sink 资源可以认为是无限制的，因而不需要担心其计算能力，同时本文考虑到信道噪声的影响，假设生

理数据包成功传输到 Sink 的概率为 $\lambda_i \in (0, 1]$ ，在每次传输之后，Sink 将 ACK 发送到传感节点 i 并没有错误发生。设 k 为帧的索引， $u_i(k) \in \{0, 1\}$ 为传感节点 i 在帧 k 处的卸载模式。若 $u_i(k) = 1$ ，则传感节点 i 进行本地卸载及状态更新；否则，传感节点 i 在 Sink 进行卸载和状态更新。设 (k, t) 为时隙的索引，其中 $t \in \{1, 2, \dots, T\}$ 。用 $d_k(t) \in \{0, 1, \dots, M\}$ 表示时隙 (k, t) 处的调度策略，其中， $d_k(t) = 0$ 表示没有调度， $d_k(t) = i$ 表示调度时隙 (k, t) 中的传感节点 i 。设 $p_{i,k}^u(t) \in \{1, 2, \dots, L_i\}$ 表示在时隙 (k, t) 开始处以卸载模式 $u_i(k)$ 传输剩余生理数据包的数量，并将 $p_{i,k}^u(t+1)$ 表示为在帧 k 结束时具有卸载模型 $u_i(k)$ 的未传输的生理数据包数量。其中， $p_{i,k}^u(t)$ 在本地和 Sink 的卸载模式是不同的。对于本地卸载模式，由于传感节点 i 在计算结束前不能被调度来传输生理数据包，因此，本地卸载模式的 $p_{i,k}^u(t)$ 的值被更新为

$$p_{i,k}^1(t+1) = \begin{cases} 0, t < c_i \text{ 或 } d_k(t) = i \text{ 且在 } t \text{ 传输成功} \\ 1, t = c_i \\ p_{i,k}^1(t), t = c_i \text{ 且 } d_k(t) \neq i \text{ 或 } t > c_i, d_k(t) = i \text{ 且在 } t \text{ 传输失败} \end{cases} \quad (1)$$

其中, c_i 是正整数, 表示传感节点 i 完成本地计算的生理数据包更新所需的时隙数量。对于在 Sink 的计算模式, 可以在每个时隙中调度一个未发送的生理数据包, 如果传感节点 i 在时隙 (k, t) 被调度并成功地传输到 Sink, 则从时隙 $(k, t + 1)$ 开始, 生理数据包的数量将减少 1, 否则将保持不变。 $p_{i,k}^u(t)$ 可表示为

$$p_{i,k}^0(t+1) = \begin{cases} p_{i,k}^0(t) - 1, & d_k(t) = i \text{ 且在 } t \text{ 成功传输} \\ p_{i,k}^0(t), & \text{其他情况} \end{cases} \quad (2)$$

传感节点 i 的能耗包括本地计算的能耗和传输生理数据包的能耗, 假设用于本地计算和传输生理数据包的能耗是恒定的, 设 e_i^c 和 e_i^r 分别表示传感节点 i 的计算功率和发射功率。当 $u_i(k) = 1$ 时, 传感节点 i 在帧 k 中进行本地计算的能耗为 $c_i e_i^c$, 否则为 0。在时隙 (k, t) 调度传感节点 i 时, 即 $d_k(t) = i$, 则在时隙 (k, t) 中的传输能耗是 e_i^r , 否则为 0。与此同时, 本文使用 AoI 来量化生理数据包的新鲜度, 新鲜度表示生理数据包生成到 Sink 接收最新更新以来的时间。设 A_i 为传感节点 i 到 Sink 时的 AoI, 如果传感节点 i 在一帧中成功发送状态更新, 则从下一帧开始时, A_i 将更新为 T ; 否则, A_i 将增加 T 。设 $g_i(k)$ 表示 Sink 从传感节点 i 接收到最新完整的状态更新时的帧数量, 则 $g_i(k)$ 的值可以更新为

$$g_i(k+1) = \begin{cases} 1, & p_{i,k}^u(T+1) = 0 \\ g_i(k) + 1, & \text{其他情况} \end{cases} \quad (3)$$

1.2 优化问题

本文的目标是通过制定卸载和调度策略来最小化 AoI 和能耗的平均加权和。设 Π_u 和 Π_d 分别表示卸载模型和调度策略。将 $\pi = (\pi_u, \pi_d)$ 表示为可行的联合卸载模式和调度策略, 其中, $\pi_u \in \Pi_u$ 和 $\pi_d \in \Pi_d$ 。在帧 k 中, 策略 π_u 需要为每个传感节点 i 设置卸载判决 $u_i(k)$; 在时隙 (k, t) 中, 策略 π_d 可以空闲或选择具有剩余生理数据包的传感节点 i 进行传输。令 $r_i(k, t)$ 表示传感节点 i 是否在时隙 (k, t) 中调度。如果传感节点 i 在时隙 (k, t) 中调度, 则 $r_i(k, t) = 1$, 否则为 0。在策略 π 下, 传感节点 i 的长期平均 AoI 为

$$\bar{A}_i^\pi = \lim_{K \rightarrow \infty} \frac{T}{K} \mathbb{E}_\pi \left[\sum_{k=1}^K g_i(k) | \mathbf{g}_1 \right] \quad (4)$$

在策略 π 下, 传感节点 i 的长期平均能耗为

$$\bar{e}_i^\pi = \lim_{K \rightarrow \infty} \frac{1}{KT} \mathbb{E}_\pi \left[\sum_{k=1}^K u_i(k) c_i e_i^c + r_i(k, t) e_i^r | \mathbf{g}_1 \right] \quad (5)$$

其中, 向量 $\mathbf{g}_1 = (g_1(1), \dots, g_M(1))$ 表示 $g_i(1)$ 的初始值。通过使用 AoI 和能耗的平均加权和来评估 WBAN 系统性能, 并构建如下优化目标

$$\min_{\pi} \sum_{i=1}^M \bar{A}_i^\pi + \beta_i \bar{e}_i^\pi \quad (6)$$

其中, $\beta_i > 0$ 是传感节点 i 的平均能耗的权重系数。

2 联合卸载与调度策略

2.1 马尔可夫决策

本文使用双层离散时间 MDP 来制定最小化目标函数的卸载和调度策略。卸载决策在每一帧的开始执行, 调度决策在每个时隙的开始执行。同时由于卸载动作和调度动作执行所依赖的状态不同, 本文对其组成部分进行分别说明。

1) 策略。通过 $u(k) = (u_i(k)) \in U$ 表示帧 k 开始处的卸载动作, 其中 U 包括所有可能的卸载动作空间。调度决策 $d_k(t) \in \{0, 1, \dots, M\}$ 只有当传感节点 i 有生理数据包等待发送且数据包数量不超过剩余时隙数时才对传感节点 i 进行调度, 即 $d_k(t) = i$ 满足 $0 < p_{i,k}^u(t) \leq T + 1 - t$ 。

2) 状态。设一个足够大的有限正整数 \hat{g}_i 为 g_i 的上限, 且 $g_i \in \{1, 2, \dots, \hat{g}_i\}$ 。令 $s(k, 0) = (g_i(k))$ 表示帧 k 的初始状态, 并且 AoI 的更新过程受到计算模式、等待传输生理数据包数量和时隙索引的影响, 因此需要记录这些参数。记 $s(k, t) = (t, s(k, 0), p(k, t), u(k))$ 为时隙 (k, t) 处的系统状态, 其中, $p(k, t) = (p_{1,k}^{u_1(k)}(t), \dots, p_{M,k}^{u_M(k)}(t))$, $t \in \{1, 2, \dots, T\}$ 。另外, 将 S_0 和 S_t 分别记为 $s(k, 0)$ 和 $s(k, t)$ 的状态空间。

3) 转移概率。由于 WBAN 系统模型中帧开始、结束以及帧内的状态转移过程不同, 状态转移分为 3 种情况。

① 当从 $s(k, 0)$ 向 $s(k, 1)$ 转移时, 卸载模式中的传感节点 i 要传输的生理数据包数量已知, 因此转移是确定性的。

② 当从 $s(k, t-1)$ 到 $s(k, t)$ 转移时, 即 $t \in \{2, \dots, T\}$, 如果传感节点 i 在时隙 $(k, t-1)$ 期间没有被调度, 则

$p_{i,k}^{u_i(k)}(t)$ 根据式(1)或式(2)更新并且转移是确定性的;如果在时隙 $(k,t-1)$ 期间传感节点 i 被调度, $p_{i,k}^{u_i(k)}(t)$ 的值需要根据反馈信息确定, 其中, $\text{Pe}(p_{i,k}^{u_i(k)}(t) = p_{i,k}^{u_i(k)}(t-1) | p_{i,k}^{u_i(k)}(t-1)) = 1 - \lambda_i$ 表示失败概率, $\text{Pe}(p_{i,k}^{u_i(k)}(t) = p_{i,k}^{u_i(k)}(t-1) - 1 | p_{i,k}^{u_i(k)}(t-1)) = \lambda_i$ 表示成功概率。

③当从 $s(k,T)$ 到 $s(k+1,0)$ 转移时, 如果传感节点 i 在时隙 (k,T) 期间空闲, 则根据式(1)~式(3)更新 $g_i(k+1)$ 的值并且转移是确定性的; 如果传感节点 i 在时隙 (k,T) 期间被调度, 则 $g_i(k+1)$ 的值根据反馈信息确定, 其中, $\text{Pe}(g_i(k+1) = g_i(k) + 1 | g_i(k)) = 1 - \lambda_i$ 表示失败概率, $\text{Pe}(g_i(k+1) = 1 | g_i(k)) = \lambda_i$ 表示成功概率。

4) 成本。在 $s(k,0)$ 状态下做出的卸载决策 $u(k)$ 的成本表示为传感节点 i 的能耗加权和, 即

$$C(s(k,0), u(k)) = \sum_{i=1}^M \beta_i u_i(k) c_i e_i^l \quad (7)$$

在 $s(k,t)$ 状态下做出的调度决策 $d_k(t)$ 的成本表示为传感节点 i 的AoI和能耗的加权和, 即

$$C(s(k,t), d_k(t)) = \sum_{i=1}^M T g_i(k) + \beta_i r_i(k,t) e_i^r \quad (8)$$

设 w 表示动作向量, 其中, $w(k,0) = u(k)$, $w(k,t) = d_k(t)$ 。在策略 π 和初始状态 s 下, 总成本表示为

$$V(s, \pi) = \lim_{K \rightarrow \infty} \mathbb{E}_\pi \left[\sum_{k=1}^K \sum_{t=1}^T C(s(k,t), w(k,t)) \right] \quad (9)$$

用 π^* 表示最小化期望总成本的最优策略, $V(s)$ 表示最优期望总成本, 则

$$V(s) = \min_{\pi^*} V(s, \pi^*) \quad (10)$$

令 s' 为动作 w 执行后 s 的下一个系统状态, $W(s)$ 表示状态为 s 的可执行动作的集合, $V(s)$ 满足贝尔曼方程并可用迭代法求解, 则有

$$V(s) = \min_{w \in W(s)} C(s, w) + \mathbb{E}[V(s') | s, w] \quad (11)$$

2.2 深度强化学习策略

本节基于DRL的方法来开发卸载和调度策略。在DRL中, 使用神经网络来进行 Q 值估计, 即深度 Q 网络(DQN, deep Q-network), 其中, 神经网络将系统状态作为输入, 每个动作对应的 Q 值作为输出, 然后用基于梯度的方法训练神经网络的权值

以最小化由输出 Q 值和目标值构成的损失函数。

对于卸载神经网络, 在状态 $s(k,0)$ 下做出卸载决策 $u(k)$ 之后, 需要执行一系列调度决策以达到下一个卸载神经网络的状态 $s(k+1,0)$ 。因此, 在 $S(k,0)$ 状态下执行动作 $u(k)$ 所产生的成本包含在当前帧中执行一系列调度动作的成本, 表示为

$$C(s(k,0), u(k)) = \sum_{i=1}^M T g_i(k) + \frac{1}{T} \sum_{i=1}^M \beta_i \sum_{t=1}^T u_i(k) c_i e_i^l + r_i(k,t) e_i^r \quad (12)$$

令 $Q(s, w; \theta)$ 表示具有权重向量 θ 的神经网络的动作价值函数。对于具有任意组合的 $w_o(k) \triangleq (s(k,0), u(k), c(s(k,0)), u(k), s(k+1,0))$, 定义卸载神经网络的损失函数为

$$L(\theta_o) = (y_o - Q(s(k,0), u(k); \theta_o))^2 \quad (13)$$

其中, θ_o 是卸载神经网络的权重向量, $y_o = c(s(k,0), u(k) + \gamma \min_{u(k+1)} Q(s(k+1,0), u(k+1); \theta_o))$ 是目标值, $\gamma \in \{0, 1\}$ 是折扣率。则关于 θ_o 的损失函数的梯度为

$$\nabla_{\theta_o} L(\theta_o) = (y_o - Q(s(k,0), u(k); \theta_o)) \cdot \nabla_{\theta_o} Q(s(k,0), u(k); \theta_o) \quad (14)$$

卸载神经网络的权值 θ_o 可以通过梯度下降法来更新。对于调度神经网络, 设 $C(s(k,t), d_k(t))$ 是在 $s(k,t)$ 状态下执行动作 $d_k(t)$ 产生的成本, 其中 $t \in \{1, \dots, T\}$ 。当 $1 \leq t \leq T-1$ 时, $s(k,t)$ 的下一个调度状态为 $s(k,t+1)$; 当 $t = T$ 时, $s(k,t)$ 的下一个调度状态为 $s(k+1,1)$ 。令神经网络中 $s(k,t)$ 的下一个调度状态为 $\bar{s}(k,t)$, 则在 $1 \leq t \leq T$ 下, $w_s(k,t) \triangleq (s(k,t), d_k(t), c(s(k,t)), d_k(t), \bar{s}(k,t))$ 的任何组合可以定义调度神经网络的损失函数为

$$L(\theta_s) = (y_s - Q(s(k,t), d_k(t); \theta_s))^2 \quad (15)$$

其中, θ_s 是调度神经网络的权重向量。

对于 $1 \leq t \leq T-1$, 目标值 $y_s = C(s(k,t), d_k(t) + \gamma \min_{d_k(t+1)} Q(s(k,t+1), d_k(t+1); \theta_s))$; 当 $t = T$ 时, $y_s = C(s(k,t), d_k(t) + \gamma \min_{d_{k+1}(1)} Q(s(k,t+1), d_{k+1}(1); \theta_s))$ 。调度神经网络关于 θ_s 的损失函数的梯度表示为

$$\nabla_{\theta_s} L(\theta_s) = (y_s - Q(s(k,t), d_k(t); \theta_s)) \cdot \nabla_{\theta_s} Q(s(k,t), d_k(t); \theta_s) \quad (16)$$

在DQN中通过使用经验回放来更新权重,即保存每次与环境交互的成本和更新状态,然后选择一批数值进行目标值更新来改善DRL策略的收敛性能。由于在执行动作之后不能立即获得卸载神经网络的成本和下一状态,本文使用后期经验存储方法,即策略不存储动作执行后的经验,而是存储系统状态转换到下一步的神经网络状态并获得所有相关成本,图2显示了异质神经网络的同步训练框架。

算法1给出了DRL策略的训练过程,同时在动作的选择上使用 ϵ -贪婪策略按概率随机选择动作来平衡探索和开发之间的权衡,否则贪婪地选择Q值最小的动作,同时考虑 $\epsilon \rightarrow 0$ 作为算法的迭代过程。

算法1 DRL策略的训练过程

- 1) 初始化: Ω_o 、 Ω_s 、 θ_o 和 θ_s ;
- 2) 循环
- 3) for $k = 1, \dots, K$
- 4) for $t = 0, \dots, T$
- 5) if $t = 0$ then
- 6) 计算卸载成本 $C(s(k-1,0), u(k-1))$ 并存储 $w_o(k-1)$ 在 Ω_o 中;
- 7) 从 Ω_o 随机抽样一批经验组合;
- 8) 计算经验 y_o 值并根据式(13)和式(14)

- 更新 θ_o ;
- 9) 基于 ϵ -贪婪策略选择 $u(k)$ 并执行动作观察成本;
- 10) else
- 11) if $t = 1$ then
- 12) $w_s(k-1, T)$ 在 Ω_s 中并从 Ω_s 随机抽样一批经验组合;
- 13) 计算经验 y_s 值并根据式(15)和式(16)更新 θ_s ;
- 14) end if
- 15) 基于 ϵ -贪婪策略选择 $d_k(t)$ 并执行动作观察成本和下一步状态;
- 16) if $1 \leq t < T$ then
- 17) 存储 $w_s(k, t)$ 在 Ω_s 中并随机抽样一批经验组合;
- 18) 计算经验 y_s 值并根据式(15)和式(16)更新 θ_s ;
- 19) end if
- 20) end if
- 21) end for
- 22) end for

3 仿真结果与分析

本文使用Python 3.8仿真验证所提MDP策略和DRL策略。其中, DRL策略通过TensorFlow实现ANN, 卸载ANN和调度ANN采用双隐层神经网络, 每个隐层包含50个REU激活神经元。同时, 2个

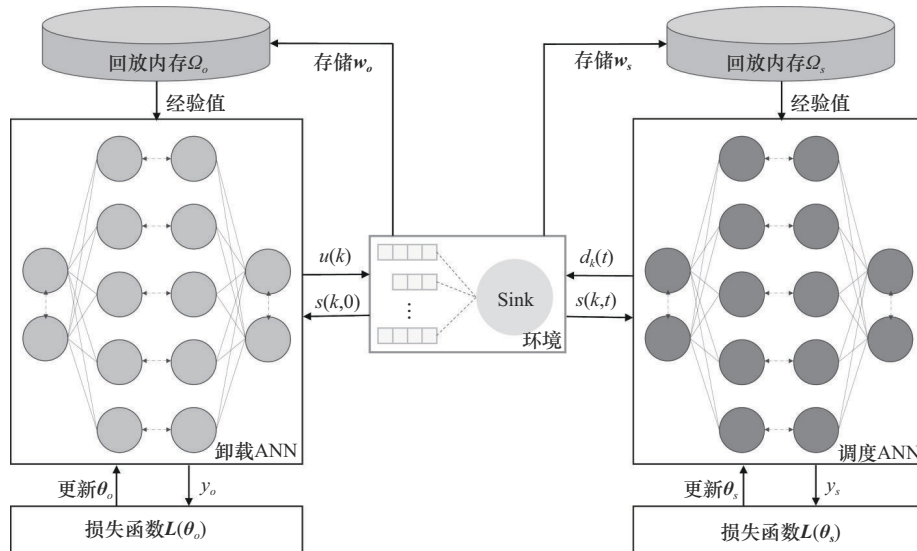


图2 同步训练框架

ANN 设置相同的超参数, 所有策略通过运行 10^5 帧获取, 如表 1 所示。此外, 本文评估了 3 种卸载策略和一种贪婪调度策略的性能。对于贪婪调度策略, 选择最大化 AoI 减去每个时隙中的加权能耗的传感节点进行调度, 同时考虑 2 种在线联合算法, 即 EG 策略和 RRG 策略。

表 1 仿真参数	
参数	值
接收更新帧数量 $g_i(1)$	1
计算功率 e_i^l/mW	0.3
传输功率 e_i^t/mW	1
每帧时隙数 T	4~8
权重因子 β_i	$\{0.5, 10, \dots, 30\}$
传感节点 i	$\{2, 3, \dots, 6\}$
损失因子 p	0.95
学习率 r	0.0001
初始状态 ε	0.5
帧数量 K	1×10^6
匹配大小	32
回放内存大小	2 000

图 3 和图 4 显示了由 MDP 策略和 DRL 策略产生的随着权重因子 β_i 增加导致的平均能耗和平均 AoI 的变化。参数设置为权重因子 $\beta_i \in \{0.5, 10, \dots, 30\}$ 、 $T = 4$ 、 $\lambda_i = 0.8$ 、 $\hat{g}_i = 10$ 、 $L_i = 3$ 以及 $c_i = i$ 。在 2 种策略下, 平均能耗随着权重因子 β_i 的增加而减少, 平均 AoI 随着权重因子 β_i 的增加而增加。这表明频繁地对传感节点进行调度可以降低系统的 AoI, 但也意味着更多的能耗, 因此在 AoI 和能耗之间存在折中。

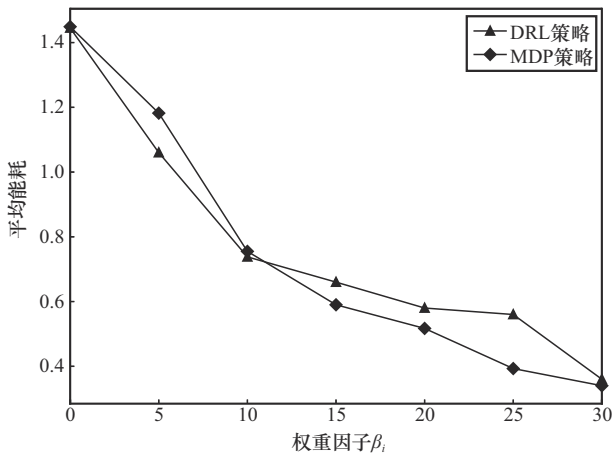


图 3 权重因子对平均能耗的影响

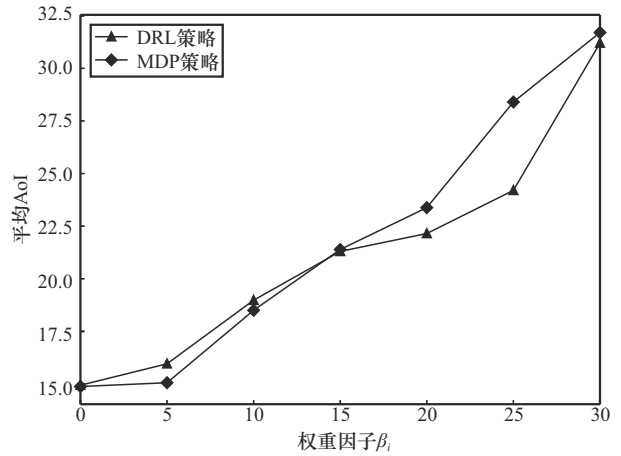


图 4 权重因子对平均 AoI 的影响

图 5 比较了 WBAN 系统在不同权重因子 $\beta_i \in \{0.5, 10, \dots, 30\}$ 下各策略的 AoI 和能耗平均加权和, 参数设置为 $T = 4$ 、 $\lambda_i = 0.8$ 、 $\hat{g}_i = 10$ 、 $L_i = 3$ 和 $c_i = i$ 。结果表明, DRL 策略的性能接近 MDP 策略并优于其他策略。此外, 在所有策略中, AoI 和能耗平均加权和随着权重因子 β_i 的增加而增加。这是因为 β_i 的增加意味着能耗权重提升而 AoI 权重保持不变, 从而导致加权和增加。

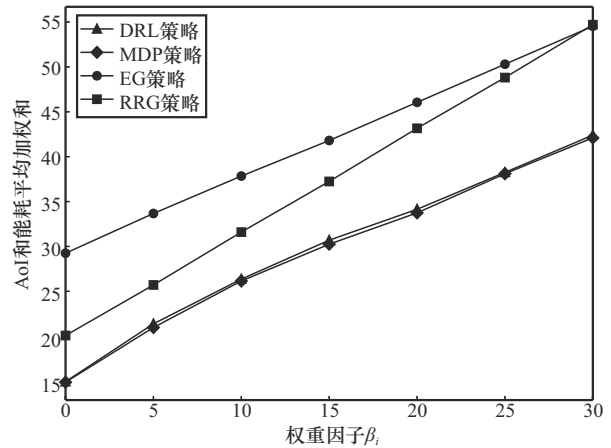


图 5 权重因子对 AoI 和能耗平均加权和的影响

图 6 评估了不同帧长 (每帧时隙数) 下各策略的性能, 参数设置为 $T \in \{4, 5, \dots, 8\}$ 、 $\beta_i = 40$ 、 $c_1 = 1$ 、 $c_2 = c_3 = 3$ 、 $L_1 = L_3 = 2$ 、 $L_2 = 3$ 、 $\hat{g}_i = 20$ 、 $\lambda_1 = 0.7$ 、 $\lambda_2 = 0.8$ 和 $\lambda_3 = 0.9$ 。由于 DRL 策略充分考虑了帧长对决策的影响, 其性能优于在线策略并与 MDP 策略性能接近。此外, MDP 策略和 DRL 策略的结果并不随着帧长的增加而单调变化。这是因为虽然帧长的增加提高了每个传感节点被调度的机会, 但也减少了更新生成的频率。

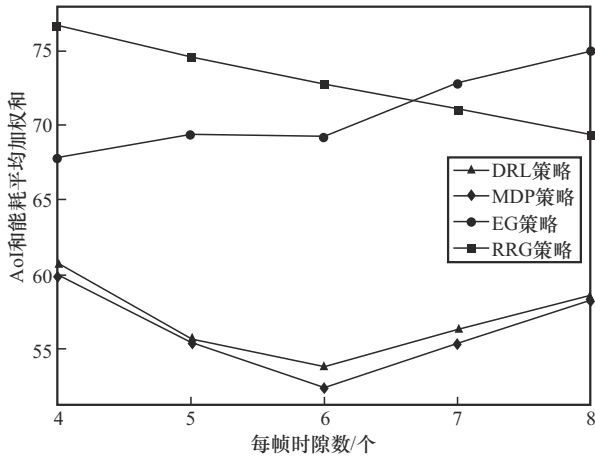


图6 每帧时隙数对AoI和能耗平均加权和的影响

图7展示了随着传感节点数量增加,各策略AoI和能耗平均加权和的变化情况,其中, $M \in \{2,3,\dots,6\}$, $T = 7$, $\beta_i = 30$, $\hat{g}_i = 30$, λ_i 从0.5~0.9中随机选取, c_i 和 L_i 从 $\{2,3,4,5\}$ 中随机选取。MDP策略的高度复杂性使其随着系统规模的增大而难以求解,而DRL策略在性能上优于其他策略。随着传感节点数量的增加,AoI和能耗平均加权和也增加,这是因为传感节点数量的增加减少了每个传感节点被调度的机会,从而导致平均AoI的值更大。

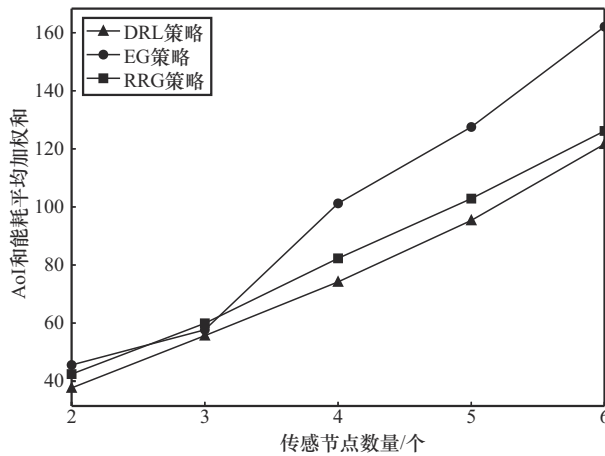


图7 传感节点对AoI和能耗平均加权和的影响

图8显示了在 $M = 3$ 、 $T = 4$ 、 $\lambda_i = 0.8$ 、 $\hat{g}_i = 10$ 、 $L_i = 3$ 、 $c_i = i$ 和 $\beta_i = 10$ 的条件下,MDP策略和DRL策略下AoI和能耗平均加权和的收敛性能。从图8中可以看出,DRL策略在训练约380s后开始趋于收敛,而MDP策略则需要约800s才能达到收敛。这主要是因为MDP策略的状态空间非

常庞大,需要花费大量时间来迭代计算每个状态的值函数。此外,由于系统具有多数据包和多时隙的特点,MDP策略在早期训练阶段优先选择了能耗最小的卸载模式和调度策略,使图8中MDP策略下初始时间段的加权和值保持不变。MDP策略需要通过迭代计算 $2^M T \prod_{i=1}^M \hat{g}_i (L_i + 1)$ 个状态值。当传感节点数量较多时,由于维数灾难的影响,MDP策略将变得难以实现。相比之下,DRL策略适用于大规模系统,通过神经网络逼近函数值,从而有效减轻维数灾难的问题。仿真结果表明,DRL策略在系统中的表现优于MDP策略,且收敛速度更快。

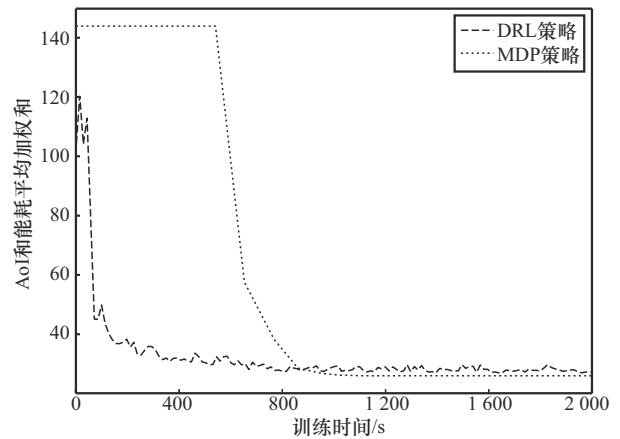


图8 不同策略下AoI和能耗平均加权和的收敛性能

4 结束语

本文研究了一种在WBAN系统中信息年龄和能耗联合优化的卸载和调度策略。通过考虑最小化AoI和能耗平均加权和来建立卸载和调度模型,并将这一问题形式化为一种离散时间马尔可夫决策过程来解决,同时本文设计了一种基于深度强化学习的策略以降低维数灾难和计算复杂度。仿真结果表明,所提策略表现出良好的性能。在未来的研究中,笔者将进一步探讨具有时变通道的卸载和调度策略以适应更为复杂的通信环境。

参考文献:

[1] JAVADPOUR A, SANGAIAH A K, JA'FARI F, et al. Toward a secure industrial wireless body area network focusing MAC layer protocols: an analytical review[J]. IEEE Transactions on Industrial Informatics, 2023, 19(2): 2028-2038.

[2] CAVALLARI R, MARTELLI F, ROSINI R, et al. A survey on wireless

- body area networks: technologies and design challenges[J]. IEEE Communications Surveys & Tutorials, 2014, 16(3): 1635-1657.
- [3] LIN Z Q, LIU X X, ZHOU H, et al. Adaptive time-varying routing for energy saving and load balancing in wireless body area networks[J]. IEEE Transactions on Mobile Computing, 2024, 23(1): 90-101.
- [4] VERMA S, KAWAMOTO Y, FADLULLAH Z M, et al. A survey on network methodologies for real-time analytics of massive IoT data and open research issues[J]. IEEE Communications Surveys & Tutorials, 2017, 19(3): 1457-1477.
- [5] KAHRAMAN İ, KÖSE A, KOCA M, et al. Age of information in Internet of Things: a survey[J]. IEEE Internet of Things Journal, 2024, 11(6): 9896-9914.
- [6] COSTA M, CODREANU M, EPHEMIDES A. On the age of information in status update systems with packet management[J]. IEEE Transactions on Information Theory, 2016, 62(4): 1897-1910.
- [7] KAUL S, YATES R, GRUTESER M. Real-time status: how often should one update? [C]//Proceedings of IEEE INFOCOM. Piscataway: IEEE Press, 2012: 2731-2735.
- [8] TANG H Y, WANG J T, SONG L Q, et al. Minimizing age of information with power constraints: multi-user opportunistic scheduling in multi-state time-varying channels[J]. IEEE Journal on Selected Areas in Communications, 2020, 38(5): 854-868.
- [9] YAO B K, GAO H, ZHANG Y, et al. Maximum AoI minimization for target monitoring in battery-free wireless sensor networks[J]. IEEE Transactions on Mobile Computing, 2023, 22(8): 4754-4772.
- [10] FOUNTOULAKIS E, CHARALAMBOUS T, EPHEMIDES A, et al. Scheduling policies for AoI minimization with timely throughput constraints[J]. IEEE Transactions on Communications, 2023, 71(7): 3905-3917.
- [11] ZAKERI A, MOLTAFFET M, LEINONEN M, et al. Minimizing the AoI in resource-constrained multi-source relaying systems: dynamic and learning-based scheduling[J]. IEEE Transactions on Wireless Communications, 2024, 23(1): 450-466.
- [12] KADOTA I, SINHA A, UYSAL-BIYIKOGLU E, et al. Scheduling policies for minimizing age of information in broadcast wireless networks[J]. IEEE/ACM Transactions on Networking, 2018, 26(6): 2637-2650.
- [13] LI J, ZHOU Y, CHEN H. Age of information for multicast transmission with fixed and random deadlines in IoT systems[J]. IEEE Internet of Things Journal, 2020, 7(9): 8178-8191.
- [14] KUANG Q B, GONG J, CHEN X, et al. Analysis on computation-intensive status update in mobile edge computing[J]. IEEE Transactions on Vehicular Technology, 2020, 69(4): 4353-4366.
- [15] SONG X X, QIN X Q, TAO Y Z, et al. Age based task scheduling and computation offloading in mobile-edge computing systems[C]//Proceedings of the 2019 IEEE Wireless Communications and Networking Conference Workshop (WCNCW). Piscataway: IEEE Press, 2019: 1-6.
- [16] SAMANTA A, MISRA S. Dynamic connectivity establishment and cooperative scheduling for QoS-aware wireless body area networks[J]. IEEE Transactions on Mobile Computing, 2018, 17(12): 2775-2788.
- [17] KIM B S, SHAH B, KIM K I. Adaptive scheduling and power control for multi-objective optimization in IEEE 802.15.6 based personalized wireless body area networks[J]. IEEE Transactions on Mobile Computing, 2023, 22(11): 6251-6268.
- [18] YI C Y, CAI J. Transmission management of delay-sensitive medical packets in beyond wireless body area networks: a queuing game approach[J]. IEEE Transactions on Mobile Computing, 2018, 17(9): 2209-2222.
- [19] LIANG B W, LIU X X, ZHOU H, et al. Channel resource scheduling for stringent demand of emergency data transmission in WBANs[J]. IEEE Transactions on Wireless Communications, 2021, 20(4): 2341-2352.
- [20] ZHANG G Y, SHEN C, SHI Q J, et al. AoI minimization for WSN data collection with periodic updating scheme[J]. IEEE Transactions on Wireless Communications, 2023, 22(1): 32-46.

[作者简介]



张政 (1989-), 男, 湖南岳阳人, 重庆邮电大学博士生, 主要研究方向为无线体域网、智慧医疗、实时调度等。



谢鑫 (1990-), 男, 河南南阳人, 博士, 重庆邮电大学讲师, 主要研究方向为工业物联网、实时调度、信息年龄优化等。



柏桐 (1987-), 男, 重庆人, 博士, 重庆邮电大学讲师, 主要研究方向为微弱信号检测、无线体域网等。



林金朝 (1966-), 男, 四川蓬溪人, 博士, 重庆邮电大学教授、博士生导师, 主要研究方向为无线通信传输、无线体域网与信息处理等。



李章勇 (1976-), 男, 重庆人, 博士, 重庆邮电大学教授, 主要研究方向为无线体域网、人机耦合、图像处理等。