基于信息年龄的工业无线传感器网络混合数据调度方法

王 恒*① 余 蕾① 谢 鑫②

①(重庆邮电大学工业物联网与网络化控制教育部重点实验室 重庆 400065)

②(重庆邮电大学通信与信息工程学院 重庆 400065)

摘 要：在工业无线传感器网络(IWSN)中，实时交付工业现场的周期性控制/传感数据流与非周期性事件数据流，是保障生产安全高效运行的关键。信息年龄(AoI)作为一种新兴的数据新鲜度衡量指标，能够从目标节点角度全面地度量网络数据交付的实时性。针对周期性和非周期性数据混合的工业无线传感器网络，该文在引入网络数据整体新鲜度指标的同时，考虑到周期性数据新鲜度在超过阈值后可能会对工业生产造成负面影响，建立了最小化系统平均AoI和周期性数据AoI逾期概率的联合优化模型，并将优化问题表述为马尔可夫决策过程(MDP)进行求解。由于传统基于相对值迭代的最优求解方法在大规模网络中因为维度灾难难以实施，因此采用深度强化学习(DRL)降低优化问题的状态空间维度，并改进决策探索机制以加快学习速度，提出了基于优化决策探索的深度强化学习(DRL-ODE)调度方法。仿真结果表明，所提方法能够提高网络数据交付的实时性，并有效减少周期性数据的AoI逾期概率。

关键词：网络调度；工业无线传感器网络；信息年龄

中图分类号：TN929.5 **文献标识码**：A **文章编号**：1009-5896(2023)03-1065-09

DOI: 10.11999/JEIT220088

Hybrid Data Scheduling Method for Industrial Wireless Sensor Networks Based on Age of Information

WANG Heng① YU Lei① XIE Xin②

①(*Key Laboratory of Industrial Internet of Things and Networked Control*, *Ministry of Education*, *Chongqing University of Posts and Telecommunications*, *Chongqing* 400065, *China*)

②(*School of Communication and Information Engineering*, *Chongqing University of Posts and Telecommunications*, *Chongqing* 400065, *China*)

Abstract: In Industrial Wireless Sensor Networks (IWSN), timely delivery of periodic control/sensing data flows and aperiodic event data flows is crucial to ensure production safety and efficiency. As a new metric of data freshness, Age of Information (AoI) can comprehensively measure the real-time performance of data delivery from the perspective of destination node. For industrial wireless sensor networks with hybrid periodic and aperiodic data, the data freshness metric of whole network is introduced. Considering that the freshness of periodic data exceeding the threshold may have a negative impact on industrial production, a joint optimization model is established, which minimizes the system average AoI and the probability of AoI overdue for periodic data, and then the optimization problem is formulated as a Markov Decision Process (MDP). Since the traditional optimal solution method based on relative value iteration is difficult to implement in large-scale networks result from dimensional disasters, Deep Reinforcement Learning (DRL) is used to reduce the state space dimension of the optimization problem. Moreover, the decision exploration mechanism is improved to speed up the learning speed, and a scheduling method of deep reinforcement learning based on Optimal Decision Exploration (DRL-ODE) is proposed. Simulation results show that the proposed method can improve the

收稿日期：2022-01-19；改回日期：2022-04-21；网络出版：2022-04-26

*通信作者： 王恒 wangheng@cqupt.edu.cn

基金项目：国家自然科学基金(61972061)，重庆市自然科学基金杰出青年基金(cstc2019jcyjjqX0012)，重庆基础研究与前沿探索项目(cstc2021ycjh-bgzxm0017)

Foundation Items: The National Natural Science Foundation of China (61972061), The Natural Science Foundation of Chongqing, for Distinguished Young Scholars (cstc2019jcyjjqX0012), The Fundamental Research and Frontier Exploration of Chongqing (cstc2021ycjh-bgzxm0017)

timeliness of network data delivery while reducing the probability of AoI overdue for periodic data effectively.

Key words: Network scheduling; Industrial Wireless Sensor Networks (IWSN); Age of Information(AoI)

1 引言

随着工业无线传感器网络(Industrial Wireless Sensor Network, IWSN)技术的不断发展，IWSN已经在汽车制造、化工生产、制药加工等工业生产场景中得到广泛应用[1,2]。在这些生产场景中，传感器需要将工业现场采集的各类数据实时地发送至控制中心，以确保控制中心能够对生产过程进行实时检测监控和精准决策控制[3,4]。因此，如何提高IWSN数据交付的实时性，成为保障工业生产安全高效的关键。近年来，一种新兴的网络实时性度量指标——信息年龄(Age of Information, AoI)被提出[5]，该指标可以从目标节点的角度对网络交付数据的新鲜度进行衡量。AoI定义为当前时间与目标节点接收到的最新数据产生时间的时间差。假设目标节点接收到的最新数据产生时间为t，那么在当前时刻k，目标节点处的AoI为$k-t$。相较于时延和交付间隔等传统衡量数据时效性的指标，AoI综合考虑了数据的生成时间和传输延迟，能够更加全面地度量IWSN数据交付的实时性与新鲜度。

AoI最初是由Kaul等人[5,6]在单源-服务器系统中为描述源数据生成状态新旧时引入的。此后，研究人员在各种网络模型下考虑了数据更新方式、信道状态以及能耗等因素，对AoI相关的优化问题展开了广泛研究。Kuang等人[7,8]聚焦于单源单目的地的边缘计算场景研究了基于AoI数据的更新策略，并推导了指数分布计算时间的封闭平均AoI。Kadota等人[9]针对周期性数据更新的多源无线广播网络，研究了噪声信道下的AoI优化问题，并设计了随机策略、Max-Weight策略和Whittle's Index策略3种低复杂度的调度方案。在定期生成数据包的物联网系统中，Yin等人[10]在相关信息源存在的条件下研究了优化AoI的调度问题，并提出了基于深度强化学习(Deep Reinforcement Learning, DRL)的调度策略。而在非周期数据更新模型的研究工作中，Hsu等人[11]将最小化长期平均AoI的优化问题表述为马尔可夫决策过程(Markov Decision Process, MDP)，提出了平稳切换型调度策略。Tang等人[12]在随机生成的工业物联网模型中，研究了带宽约束下的AoI优化问题，并开发了渐近最优截断调度策略。与文献[9–12]中所研究的单信道调度问题不同，文献[13]考虑了多信道网络中的信道冲突和链路冲突，将AoI的优化问题转化为李雅普诺夫优化问题，提出了最小化AoI的链路调度方法。文献[14]在多信道系统中研究了吞吐量不损失条件下的AoI优化问题，提出了抢占式优先服务策略。此外，考虑到网络能耗因素，文献[15]在能量资源约束的条件下提出了基于Whittle Index的调度算法，而文献[16]则以最小化AoI和能耗的平均加权和为目标，提出了基于DRL的调度策略。

上述AoI调度方法均是基于单一的数据生成模型进行研究。在IWSN场景中，既有非周期生成的事件数据流，又有周期生成的控制/传感数据流[17,18]，这两类数据都需要实时地更新至控制中心，以对整个工业生产过程进行实时的监控或决策。在此类混合数据更新方式下，文献[19]在周期性和随机性两种数据生成方式混合的场景中研究了多时隙的链路调度问题，并提出了最小化平均AoI的李雅普诺夫漂移策略。文献[20]则考虑了任意采样、每时隙采样和周期采样3种数据生成方式的混合，并设计了一个低复杂度的调度算法来最小化平均加权AoI。虽然混合数据更新方式下的平均AoI调度问题已经得到了初步研究，但是现有的工作尚未考虑混合更新下周期性数据瞬时AoI对系统性能的影响。在IWSN中，周期性数据通常与控制任务相关，对信息的时效性有更高的要求[18,21]。例如，在面向工业过程自动化的工业无线网络(Wireless networks for Industrial Automation-Process Automation, WIA-PA)中，周期性控制/传感数据用于精准高效的控制任务，非周期性事件数据用于工厂的观测任务[22]。若控制中心可用的控制/传感数据瞬时AoI超过某个阈值，数据所表达的工业现场状态过于陈旧，有可能导致错误的决策，甚至影响整个系统的运行；而用于记录工业运行过程的非周期性事件数据，更关注于尽可能提高数据的新鲜度，通常没有严格的时效性约束[12]。因此，如何在混合更新的IWSN中，提高网络数据交付实时性的同时，减小周期性数据瞬时AoI超过阈值时的逾期概率，具有重要的研究意义。

针对上述问题，本文基于周期/非周期性数据更新方式混合并存的IWSN，研究了最小化系统平均AoI与周期性数据AoI逾期概率加权和的优化问题，并将该多目标优化问题表述为MDP进行求解。同时，为避免传统最优求解方法可能会遭受的维数灾难问题，采用了DRL方法在学习最优策略的同时降低状态空间维度。此外，为了有效加快学习速度，本文对传统DRL方法的探索策略进行改进，

提出了基于优化决策探索的深度强化学习(Deep Reinforcement Learning based on Optimal Decision Exploration, DRL-ODE)调度方法。仿真结果验证了所提调度方法的有效性。

2　系统模型与优化问题

2.1　系统模型

考虑一个由M个源节点(从站)和单个目标节点(主站或控制中心)组成的IWSN，如图1所示。定义$\Phi=\phi_1\cup\phi_2$为网络中M个源节点构成的集合，其中集合ϕ_1的源节点采集非周期性生成的事件数据，集合ϕ_2的源节点采集周期性生成的控制/传感数据。被采集的数据将通过有限的信道资源传输至目标节点。

在非周期性数据和周期性数据更新方式混合并存的情况下，假设系统调度更新过程是基于时隙的，并且源节点$m\in\Phi$的一个数据在一个时隙中传输[3]，令$k\in\{1,2,\cdots,K\}$表示时隙的索引。在时隙k进行数据交付时，由于信道噪声的存在，源节点m的传输成功率为$q_m\in(0,1]$。若数据交付成功，源节点会根据目标节点回复的确认帧将存储队列中已发送过的数据丢弃；若数据交付失败，则目标节点会请求重传。在源节点的有限存储队列中，源节点按先入先出的方式缓存新采集的数据。当存储队列满载时，若有新数据到来，则将队列头部数据丢弃并把新数据缓存至队列尾部。

2.2　问题建模

为了有效度量网络调度过程中数据混合更新至目标节点的实时性，本文使用AoI来衡量数据在目标节点处的新鲜度。在时隙k进行调度时，由于信道资源的限制，网络需做出合适的调度决策选择源节点完成数据交付。假设调度决策为$d(k)$，则$d(k)\in\{0,1,\cdots,M\}$，其中$d(k)=0$表示网络处于空闲状态。若$d(k)\neq0$，目标节点将调度$d(k)$对应源节点m存储队列中的头部数据。定义$u_m(k)\in\{0,1\}$表示是否成功交付数据，在决策执行后，若成功交付数据$u_m(k)=1$，新数据将会替代旧数据；若交付失败$u_m(k)=0$，旧数据的AoI自加1。因此，在完成1次决策后目标节点处数据AoI $a_m(k)$的更新过程为

$$a_m(k+1)=\begin{cases}z_m^h(k)+1, & u_m(k)=1\\ a_m(k)+1, & u_m(k)=0\end{cases}\tag{1}$$

其中，$z_m^h(k)$表示源节点m存储队列头部数据的AoI，h为头部数据标识。

对于$m\in\phi_1$的源节点，假设其采集的非周期性数据产生方式服从$\lambda_{m\in\phi_1}\in(0,1]$的伯努利分布。定义$g_{m\in\phi_1}(k)\in\{0,1\}$表示源节点$m$在时隙$k$是否成功采集到新数据，如果成功采集则$g_{m\in\phi_1}(k)=1$，否则$g_{m\in\phi_1}(k)=0$。当$g_{m\in\phi_1}(k)=1$时，被采集的新数据将被缓存至长度为$L_{m\in\phi_1}$的存储队列中。令$l_{m\in\phi_1}(k)\in\{0,1,\cdots,L_{m\in\phi_1}\}$为存储队列当前缓存数据个数，若存储队列$l_{m\in\phi_1}>1$，则头部数据的后继数据AoI为$z_{m\in\phi_1}^n(k)$，其中$n$为后继数据标识。由此，非周期性数据源节点的存储队列头部数据AoI$z_{m\in\phi_1}^h(k)$的迭代过程可以表述为

$$z_{m\in\phi_1}^h(k+1)=\begin{cases}1, & \text{当}l_{m\in\phi_1}(k)=0\text{且}g_{m\in\phi_1}(k)=1\\ z_{m\in\phi_1}^n(k)+1, & \text{当}l_{m\in\phi_1}(k)>1\text{且}u_m(k)=1;\text{当}l_{m\in\phi_1}(k)=L_{m\in\phi_1}\text{且}g_{m\in\phi_1}(k)=1\\ z_{m\in\phi_1}^h(k)+1, & \text{当}l_{m\in\phi_1}(k)\neq0\text{且}u_m(k)=0\end{cases}\tag{2}$$

对于$m\in\phi_2$的源节点，定义其采集的周期性数据产生周期为$T_{m\in\phi_2}$。当$k\bmod T_{m\in\phi_2}=0$时表示源节点在时隙k采集到新数据，其中$k\bmod T_{m\in\phi_2}$表示k对$T_{m\in\phi_2}$的取余操作。与非周期性数据源节点的存储队列头部数据AoI更新类似，周期性数据源节点的$z_{m\in\phi_2}^h(k)$迭代过程可以表示为

$$z_{m\in\phi_2}^h(k+1)=\begin{cases}1, & \text{当}l_{m\in\phi_2}(k)=0\text{且}k\bmod T_{m\in\phi_2}=0\\ z_{m\in\phi_2}^n(k)+1, & \text{当}l_{m\in\phi_2}(k)>1\text{且}u_m(k)=1;\text{当}l_{m\in\phi_2}(k)=L_{m\in\phi_2}\text{且}k\bmod T_{m\in\phi_2}=0\\ z_{m\in\phi_2}^h(k)+1, & \text{当}l_{m\in\phi_2}(k)\neq0\text{且}u_m(k)=0\end{cases}\tag{3}$$

源节点$m\in\phi_2$采集的周期性数据在交付时，其瞬时AoI $a_{m\in\phi_2}(k)$需要尽可能在阈值内，以保障数据所表达信息的新鲜度。本文在长期时间条件下统计目标节点所有周期性数据AoI超过其阈值的事件发生总数，并由该总数除以周期性数据源节点个数I与时隙数量K之积表示周期性数据AoI的逾期概率F，定义为

$$F=\frac{1}{IK}\sum_{k=1}^{K}\sum_{m\in\phi_2}\varepsilon\left(a_{m\in\phi_2}(k)>x_{m\in\phi_2}\right)\tag{4}$$

其中，$x_{m\in\phi_2}$代表源节点m交付数据的AoI阈值，

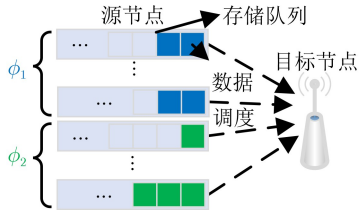


图 1 混合更新的工业无线传感器网络示意图

$\varepsilon\left(a_{m\in\phi_2}(k)>x_{m\in\phi_2}\right)\in\{0,1\}$表示$a_{m\in\phi_2}(k)>x_{m\in\phi_2}$事件发生的指示函数，如果$a_{m\in\phi_2}(k)>x_{m\in\phi_2}$事件发生，则$\varepsilon\left(a_{m\in\phi_2}(k)>x_{m\in\phi_2}\right)=1$，否则$\varepsilon\left(a_{m\in\phi_2}(k)>x_{m\in\phi_2}\right)=0$。

2.3 优化问题

为了在优化数据交付实时性的同时，减小周期性数据瞬时AoI超过阈值时的概率，本文建立了由系统平均AoI和周期性数据AoI逾期概率F组成加权和的联合优化目标函数。整个优化问题的目标函数为

$$H=\lim_{K\to\infty}\mathrm{E}\left[\frac{1}{K}\sum_{k=1}^{K}\left(\frac{1}{M}\sum_{m\in\Phi}a_m(k)+\alpha\cdot\frac{1}{I}\sum_{m\in\phi_2}\varepsilon\left(a_{m\in\phi_2}(k)>x_{m\in\phi_2}\right)\right)\right] \tag{5}$$

其中，α为系统平均AoI与逾期概率F之间的权重参数，α越大表示目标函数更关注周期性数据AoI的逾期概率。

定义集合Π表示所有的可行调度方法，则以最小化加权和H为目标的方法$\pi\in\Pi$。由此在可行的调度方法下最小化目标函数式(5)的最优方法π^*可表述为

$$\pi^*=\arg\min_{\pi\in\Pi}H_\pi \tag{6}$$

3 调度方法

为得到最小化加权和H的可行调度方法，一种潜在的最优方法是将式(6)的随机优化问题表述为MDP，然后通过相对值迭代得到调度方法[23]。然而，由于MDP方法的状态空间需考虑各个源节点的缓存信息，使得模型出现状态空间呈指数增长的情况，导致通过值迭代方式求解优化问题时计算的复杂度显著增加，以致于无法求解。相比于MDP，DRL方法具有对高维数据感知能力强的优势，能够直接从大状态空间中学习可行调度策略并解决MDP方法计算复杂度过高的问题[24]。因此，本文将开发一个基于DRL的调度方法。

3.1 基于深度强化学习的调度方法

DRL方法通过神经网络与环境交互，在最大化累积奖励值的过程中学习调度策略。其具体流程为：神经网络在当前环境状态$s(k)$下选择要执行的动作$d(k)$，然后得到下一个环境状态$s(k+1)$，在此过程中获得一个奖励值$r(k)$。在连续迭代之后，使得累积奖励值$\sum_{k=0}^{\infty}\gamma^k r(k)$最大化，其中$\gamma\in[0,1]$表示折扣因子。

与传统DRL最大化累计奖励值不同，由于式(6)的随机优化问题是最小化加权和H，当神经网络与IWSN环境进行交互时会得到一个惩罚值$c(k)$，本文的DRL方法则在最小化累计惩罚值$\sum_{k=0}^{\infty}\gamma^k c(k)$的过程中学习可行的调度策略。在学习过程中所使用的状态空间$s(k)$、动作空间$D(k)$和惩罚值函数$c(k)$具体定义如下：

(1) 系统状态空间$s(k)$包含了在时隙k时目标节点处各个源节点数据的AoI集合$a(k)$，其中$a(k)=\{a_1(k),a_2(k),\cdots,a_M(k)\}$，同时还需包含全部源节点的存储队列信息的AoI集合$z(k)$。定义$z_m(k)$表示源节点m在时隙k的缓存信息，则$z(k)=\{z_1(k),z_2(k),\cdots,z_M(k)\}$。因此，系统状态空间为

$$s(k)=(a(k),z(k)) \tag{7}$$

(2) 系统动作空间$D(k)$包括了所有可能的链路调度决策以及网络空闲时。所以，系统动作空间为

$$D(k)=\{0,1,\cdots,M\} \tag{8}$$

(3) 本文将链路调度决策后目标节点处平均AoI与周期性数据瞬时AoI超过阈值所实施的惩罚这两项加权和作为惩罚函数$c(k)$。由此，得出系统的$c(k)$为

$$c(k)=\sum_{m\in\Phi}a_m(k)-\alpha\cdot\varepsilon_m\left(a_{m\in\phi_2}(k)>x_{m\in\phi_2}\right) \tag{9}$$

其中，$\varepsilon_m\left(a_{m\in\phi_2}(k)>x_{m\in\phi_2}\right)$为决策$d(k)=m$时$a_{m\in\phi_2}(k)>x_{m\in\phi_2}$事件发生的指示函数。

在建立好神经网络与IWSN交互所需的状态空间、动作空间以及惩罚函数后，使用深度Q学习(Deep Q Network, DQN)网络训练基于DRL的调度方法[24]。图2为基于DQN网络最小化加权和H过程中学习调度策略的训练框架，网络中包含了两个结构相同但参数不同的当前值网络与目标值网络。其中，当前值网络使用最新的参数，而目标值网络使用过去的参数。在训练过程中，网络通过μ - greedy策略对IWSN环境进行探索。在决策探索时，网络会生成一个随机数$b\in[0,1]$，当$b<\mu$时当前值网络随机在动作空间中选择决策$d(k)$，否则选择最小值函数$V(s(k),d(k)|\boldsymbol{w})$对应的决策$d(k)$，值函数$V(s(k),d(k)|\boldsymbol{w})$的更新过程为

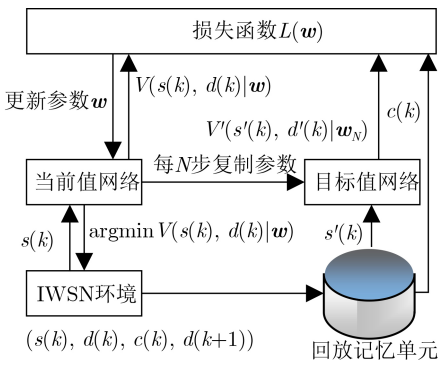


图 2 DRL调度方法训练示意图

$$V\left(s\left(k\right),d\left(k\right)|\boldsymbol{w}\right)=c\left(k\right)+\gamma\min V\left(s\left(k+1\right),d\left(k\right)|\boldsymbol{w}\right) \tag{10}$$

其中，$\boldsymbol{w}$表示当前值网络的参数向量，$c\left(k\right)$和$s\left(k+1\right)$为决策$d\left(k\right)$执行后IWSN环境返回的惩罚函数和下一状态空间。网络将返回的参数组成经验集合$\left(s\left(k\right),d\left(k\right),c\left(k\right),s\left(k+1\right)\right)$保存在回放记忆单元中。随着训练的进行，当探索因子$\mu$趋近于0时，当前值网络每次选择最小值函数$V\left(s\left(k\right),d\left(k\right)|\boldsymbol{w}\right)$对应的决策。

在学习调度策略的过程中，当前值网络的非线性函数在最小化值函数$V\left(s\left(k\right),d\left(k\right)|\boldsymbol{w}\right)$时会出现振荡的问题，所以为解决该问题每迭代$N$次当前值网络都会将自身参数向量拷贝至目标值网络。每次调度决策执行后网络会计算当前值网络与目标值网络之间的损失函数$L\left(\boldsymbol{w}\right)$，具体为

$$L\left(\boldsymbol{w}\right)=\left(V\left(s\left(k\right),d\left(k\right)|\boldsymbol{w}\right)-V'\left(s'\left(k\right),d'\left(k\right)|\boldsymbol{w}_N\right)\right)^2 \tag{11}$$

其中，$V'\left(s'\left(k\right),d'\left(k\right)|\boldsymbol{w}_N\right)$表示目标值网络的值函数，$s'\left(k\right)$和$d'\left(k\right)$分别表示从回放记忆单元中随机抽取的状态决策对，$\boldsymbol{w}_N$表示目标值网络参数向量。通过损失函数网络再进一步计算梯度损失函数，可得

$$\begin{aligned}\nabla_{\boldsymbol{w}}L\left(\boldsymbol{w}\right)=&\left(V\left(s\left(k\right),d\left(k\right)|\boldsymbol{w}\right)\right.\\&\left.-V'\left(s'\left(k\right),d'\left(k\right)|\boldsymbol{w}_N\right)\right)\\&\times\nabla_{\boldsymbol{w}}V\left(s\left(k\right),d\left(k\right)|\boldsymbol{w}\right)\end{aligned} \tag{12}$$

其中，$\nabla_{\boldsymbol{w}}$表示当前值网络的梯度向量。网络则根据式(12)采用梯度下降法更新当前值网络参数向量。随着网络迭代的进行，在损失函数趋于稳定时保存当前值网络参数向量并生成最小化加权和H的调度网络，最终得到基于DRL的链路调度方法DRL-based。

相较于MDP方法，DRL-based调度方法可克服大状态空间下MDP方法遇到的维度灾难问题。在MDP方法的设计过程中，经典的相对值迭代策略需要迭代计算有限状态空间内每一个状态的价值函数[25]。因此，在本文的优化问题中，需要限定系统的最大AoI。假设最大AoI的值为B，且所有源节点存储队列长度之和为$\sum_{m\in\Phi}L_m=L$，根据系统状态$s\left(k\right)$，需要计算B^{M+L}个状态的价值函数。随着状态数的增加，MDP方法将因为维度灾难的问题而无法实施。例如，在一个$M=4$, $B=100$且每个源节点存储队列长度为$L_{m\in\Phi}=3$的网络中，状态的数量为10^{32}个，难以迭代获得每个状态的价值函数。而DRL方法利用了深度神经网络来逼近状态的价值函数，仅需输入大小为$M+L$的状态$s\left(k\right)$特征向量，即可通过智能体与环境的交互训练逼近最优策略[24]。因此，与MDP方法相比，DRL-based调度方法更适用于本文的多传感器调度场景。

3.2　优化决策探索

上述DRL-based调度方法在最小化加权和H的训练过程中是通过传统的μ - greedy策略以概率μ从动作空间中随机选择决策进行策略探索，而在这种随机情况下低效的决策可能被选择从而影响网络学习速度。因此，本文试图通过对基于μ - greedy策略的决策探索机制进行改进以提高损失函数的收敛速度。在文献[19]中，一种启发式的Max-Ratio方法被提出用来最小化系统平均AoI，该方法会计算调度时目标节点处数据AoI的期望下降值，然后选择期望值最大的动作作为调度决策。受此启发，本文引入Max-Ratio方法评价动作空间中各个决策，并生成J个候选决策组成优势动作空间，将低效决策提前过滤。

在每个时隙探索策略前，会计算目标节点处数据$a_m\left(k\right)$与存储队列头部数据$z_m^h\left(k\right)$的差值，然后将差值与传输成功率q_m相乘得到期望下降值e_m。具体计算过程为

$$e_m=q_m\left(a_m\left(k\right)-z_m^h\left(k\right)\right) \tag{13}$$

在获得各源节点与目标节点之间的e_m后，通过e_m把所有可执行的决策做排序，有效筛选出较好的决策。然后选择e_m由大到小对应的J个决策组成优势动作空间$O\left(k\right)$，优势动作空间$O\left(k\right)$定义为

$$O\left(k\right)=\left\{o_1,o_2,\cdots,o_J\right\} \tag{14}$$

其中，$O\left(k\right)\subset D\left(k\right)$, $J<M$。在网络探索时，当前值网络从该优势动作空间中随机选择决策，然后将获取的决策与IWSN环境进行交互并学习。至此，可以获得改进的DRL-ODE调度方法，算法流程如算法1所示。

4　仿真结果与分析

本文将通过数值结果验证所提DRL-based和DRL-ODE调度方法的性能。此外，还将与文献[9]中的Greedy调度方法、文献[19]中的Max-Ratio调度方法进行对比。Greedy方法每次调度目标节点处数据AoI值最大的源节点，Max-Ratio方法调度目标节点处AoI期望下降值最大的源节点。数值结果验证平台为Windows 10操作系统、酷睿i5-9500处理器和16 GB内存的计算机。对于所提DRL-based和DRL-ODE调度方法涉及的网络，使用Python 3.5和TensorFlow 2.0软件平台训练。具体网络训练参数设置如下：当前值网络和目标值网络具有50个神经元，回放记忆单元的容量为2000，批处理大小为32，学习速率为0.001，折扣因子γ为0.9，当前值网络每100步将网络参数向量拷贝至目标值网络。

图3展示了各方法在不同网络规模下的性能表

现。其中源节点数量由2个增加到10个，每次增加1个非周期性数据源节点和1个周期性数据源节点。系统数据生成参数λ_m和T_m分别在区间[0.01, 0.2]和[10, 20]随机产生，各源节点与目标节点交付数据的传输成功率q_m取自区间[0.1, 1]。通过图3(a)可以发现，与其他调度方法相比，所提的DRL-ODE方法和DRL-based方法在各网络规模下能够对系统平均AoI表现出稳定的优化性能。进一步可以看出，随着网络规模变大，系统平均AoI也相应增加。这是由于网络中的源节点数越多，对目标节点的竞争就越激烈，使得单个源节点的更新频率变小，从而导致系统平均AoI增加。图3(b)给出了各方法周期性数据AoI逾期概率F的性能表现。从中可以看出，当网络规模较小时，各调度方法都能够有效抑制逾期概率F，这是因为源节点数量较少时信道资源竞争小，来自各个源节点的周期数据能够及时交付至目标节点。但当网络规模增大时，Max-Ratio方法和Greedy方法逾期概率F陡然变大，这是由于随着源节点数量增多，有限的信道资源导致源节点没能在阈值超过前将周期数据交付至目标节点。而DRL-ODE方法和DRL-based方法在大网络规模下依然能够有效抑制逾期概率F，这是因为所提方法在训练时学习了即将超过阈值的系统状态，提前做出了合适的决策去避免超过阈值事件的发生。综上可以看出，本文所提的方法可在各网络规模下提升网络数据传输的实时性，并保障周期数据的时效性和准确性。

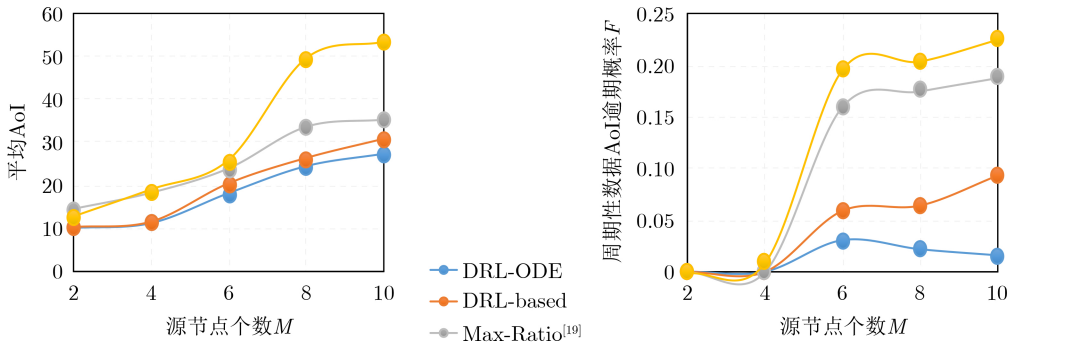
算法1 DRL-ODE调度方法训练算法流程

(1) 初始化：网络参数$\boldsymbol{w}$和$\boldsymbol{w}_N$以及回放记忆单元
(2) for $k=0,1,\cdots,K$ do
(3) 生成一个0和1之间的随机数b;
(4) if $b<\mu$ then
(5) 根据式(12)计算每个源节点的期望下降值e_m并生成优化动作空间$O(k)$;
(6) 从$O(k)$中随机选取决策$d(k)$;
(7) else then
(8) 选择$\min V(s(k),d(k)|\boldsymbol{w})$对应的决策$d(k)$;
(9) end
(10) 当前值网络执行决策$d(k)$并与当前系统状态为$s(k)$的IWSN环境交互;
(11) 获取IWSN环境反馈的下一状态$s(k+1)$和惩罚$c(k)$;
(12) 将当前经验集合$(s(k),d(k),c(k),s(k+1))$存入回放记忆单元;
(13) 从回放记忆单元中随机选取经验集合并根据式(11)计算损失函数$L(\boldsymbol{w})$;
(14) 根据式(12)利用梯度下降法更新参数向量$\boldsymbol{w}$;
(15) 每迭代N次将当前值网络参数拷贝至目标值网络;
(16) end for

不同信道质量下各调度方法的表现由图4所示。仿真中传输成功率q_m从0.5增加到0.9，并保持各个源节点至目标节点的传输成功率一致，其余相关参数设置与图3中$M=10$时保持一致。结果表明，随着传输成功率的增加，系统平均AoI逐步减小。这是因为更高的传输成功率允许更多时隙成功地交付了数据，减少了重传过程。进一步，从仿真结果中可以看出本文所提的DRL-ODE方法和DRL-based方法在各种传输成功率下都能表现出更优越的性能。

图5给出了优化决策探索后的性能改进，其中相关参数与图3中$M=10$时保持一致，每轮迭代包含10^4个时隙。从图5的仿真结果可看出，相比于DRL-based方法，DRL-ODE方法在训练开始时有更快的收敛速度，并且随着迭代的进行，DRL-ODE方法在对系统平均AoI的优化上表现出了更好的性能。这是由于在探索IWSN环境时，网络从优势动作空间中选择了较好的决策进行学习，这为后面继续学习其他状态决策对提供了一个良好的基础。因此，优化决策探索既改善了调度方法的学习效果又提升了网络的收敛速度。

图6展示了DRL-ODE方法中权重参数α的设置



(a) 不同网络规模下系统平均AoI对比 (b) 不同网络规模下周期性数据AoI逾期概率F对比

图 3 不同网络规模下各方法性能对比

对系统平均AoI和周期性数据AoI逾期概率F的影响，其中相关参数设置和图3中$M=10$时相同。从图6(a)可以看出，当α小于100时，α与系统平均AoI呈现负相关的关系；而当α大于100后，α越大系统平均AoI越大。与图6(a)展示的情况不同，图6(b)中整体趋势是α越大则逾期概率F越小，但当α增加到一定程度后，逾期概率F的变化并不明显。上述现象的出现是由于当α过小时，对F的优化权重不够，使得周期性数据超过阈值事件发生次数较多，不但导致F比较大，也造成了平均AoI相应增大；而当α过大时，因为过于关注减小F，优先交付了周期性数据，导致其他非周期性数据的新鲜度没有得到保障。综合来看，权重参数α的选择是在一个合理的区间中，过小或过大都会影响网络学习调度策略的效果。

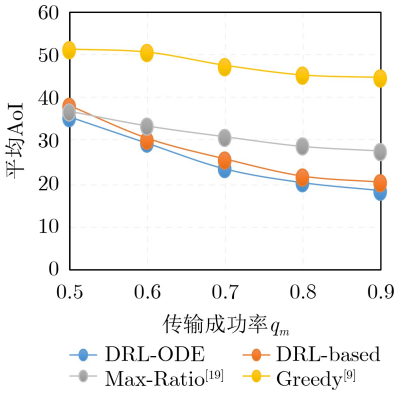


图 4 不同传输成功率下系统平均AoI对比

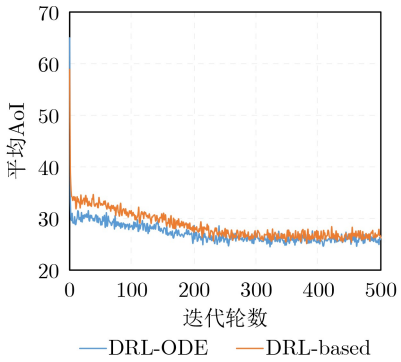
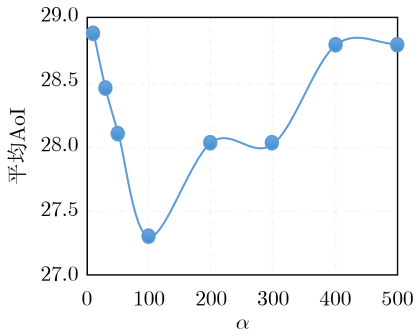
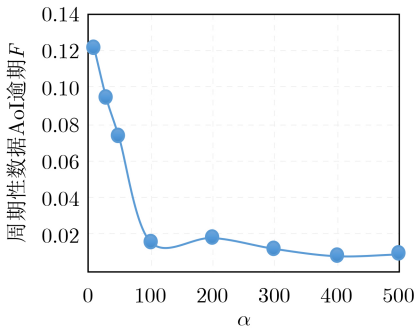


图 5 优化决策探索性能对比



(a) 不同α下系统平均AoI对比



(b) 不同α下周期性数据AoI逾期概率F对比

图 6 不同α的结果对比

5　结束语

本文针对非周期/周期两类数据更新方式混合并存的IWSN提出了基于DRL的调度方法。本方法将系统平均AoI与周期性数据AoI逾期概率加权和的优化问题表述为MDP，并利用DRL方法解决模型求解时遇到的大状态空间问题，同时为优化DRL方法的学习效果与收敛速度，结合Max-Ratio方法提出优化动作空间改进DRL方法的探索状态决策对机制。仿真结果表明，相较于其他方法，本文所提的DRL-ODE和DRL-based调度方法能够有效优化系统平均AoI和降低周期性数据瞬时AoI超过阈值的概率。在未来的研究工作中，将进一步研究AoI和能耗、丢包率等其他性能指标的联合优化问题。

参 考 文 献

[1]　胡致远, 胡文前, 李香, 等. 面向业务可达性的广域工业互联网调度算法研究[J]. 电子与信息学报, 2021, 43(9): 2608–2616. doi: 10.11999/JEIT200583.

HU Zhiyuan, HU Wenqian, LI Xiang, *et al*. Research on wide area industrial internet scheduling algorithm based on service reachability[J]. *Journal of Electronics & Information Technology*, 2021, 43(9): 2608–2616. doi: 10.11999/JEIT200583.

[2]　SHA M, GUNATILAKA D, WU Chengjie, *et al*. Empirical study and enhancements of industrial wireless sensor–actuator network protocols[J]. *IEEE Internet of Things Journal*, 2017, 4(3): 696–704. doi: 10.1109/JIOT.2017.2653362.

[3]　王恒, 朱元杰, 杨杭, 等. 基于优先级分类的工业无线网络确定性调度算法[J]. 自动化学报, 2020, 46(2): 373–384. doi: 10.16383/j.aas.c170722.

WANG Heng, ZHU Yuanjie, YANG Hang, *et al*.

Deterministic scheduling algorithm with priority classification for industrial wireless networks[J]. *Acta Automatica Sinica*, 2020, 46(2): 373–384. doi: 10.16383/j.aas.c170722.

[4] 段洁, 胡显静, 林欢, 等. 面向物联网数据特征的信息中心网络缓存方案[J]. 电子与信息学报, 2021, 43(8): 2240–2248. doi: 10.11999/JEIT200631.
DUAN Jie, HU Xianjing, LIN Huan, *et al*. Information-centric networking caching scheme for data characteristics of internet of things[J]. *Journal of Electronics & Information Technology*, 2021, 43(8): 2240–2248. doi: 10.11999/JEIT200631.

[5] KAUL S, YATES R, and GRUTESER M. Real-time status: How often should one update?[C]. 2012 Proceedings IEEE INFOCOM, Orlando, USA, 2012: 2731–2735. doi: 10.1109/INFCOM.2012.6195689.

[6] KAM C, KOMPELLA S, NGUYEN G D, *et al*. Effect of message transmission path diversity on status age[J]. *IEEE Transactions on Information Theory*, 2016, 62(3): 1360–1374. doi: 10.1109/TIT.2015.2511791.

[7] KUANG Qiaobin, GONG Jie, CHEN Xiang, *et al*. Age-of-information for computation-intensive messages in mobile edge computing[C]. The 11th International Conference on Wireless Communications and Signal Processing (WCSP), Xi'an, China, 2019: 1–6. doi: 10.1109/WCSP.2019.8927944.

[8] KUANG Qiaobin, GONG Jie, CHEN Xiang, *et al*. Analysis on computation-intensive status update in mobile edge computing[J]. *IEEE Transactions on Vehicular Technology*, 2020, 69(4): 4353–4366. doi: 10.1109/TVT.2020.2974816.

[9] 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. doi: 10.1109/TNET.2018.2873606.

[10] YIN Bo, ZHANG Shuai, and CHENG Yu. Application-oriented scheduling for optimizing the age of correlated information: A deep-reinforcement-learning-based approach[J]. *IEEE Internet of Things Journal*, 2020, 7(9): 8748–8759. doi: 10.1109/JIOT.2020.2996562.

[11] HSU Y P, MODIANO E, and DUAN Lingjie. Age of information: Design and analysis of optimal scheduling algorithms[C]. 2017 IEEE International Symposium on Information Theory (ISIT), Aachen, Germany, 2017: 561–565. doi: 10.1109/ISIT.2017.8006590.

[12] TANG Haoyue, WANG Jintao, SONG Linqi, *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. doi: 10.1109/JSAC.2020.2980911.

[13] 王恒, 段思勰, 谢鑫. 基于信息年龄优化的多信道无线网络调度方法[J]. 电子与信息学报, 2022, 44(2): 702–709. doi: 10.11999/JEIT210107.
WANG Heng, DUAN Sixie, and XIE Xin. Scheduling method for multi-channel wireless networks based on optimization of age of information[J]. *Journal of Electronics & Information Technology*, 2022, 44(2): 702–709. doi: 10.11999/JEIT210107.

[14] BEDEWY A M, SUN Yin, and SHROFF N B. Optimizing data freshness, throughput, and delay in multi-server information-update systems[C]. 2016 IEEE International Symposium on Information Theory (ISIT), Barcelona, Spain, 2016: 2569–2573. doi: 10.1109/ISIT.2016.7541763.

[15] 赵悦超, 杨涛, 胡波. 无线传感器网络中基于信息年龄的状态更新策略[J]. 微电子学与计算机, 2020, 37(11): 29–34. doi: 10.19304/j.cnki.issn1000-7180.2020.11.006.
ZHAO Yuechao, YANG Tao, and HU Bo. A status updating policy based on age of information in wireless sensor network[J]. *Microelectronics & Computer*, 2020, 37(11): 29–34. doi: 10.19304/j.cnki.issn1000-7180.2020.11.006.

[16] XIE Xin, WANG Heng, and WENG Mingjiang. A reinforcement learning approach for optimizing the age-of-computing-enabled IoT[J]. *IEEE Internet of Things Journal*, 2022, 9(4): 2778–2786. doi: 10.1109/JIOT.2021.3093156.

[17] KASHEF M and MOAYERI N. Real-time scheduling for wireless networks with random deadlines[C]. The 13th International Workshop on Factory Communication Systems (WFCS), Trondheim, Norway, 2017: 1–9. doi: 10.1109/WFCS.2017.7991954.

[18] JIN Xi, KONG Fanxin, KONG Linghe, *et al*. A hierarchical data transmission framework for industrial wireless sensor and actuator networks[J]. *IEEE Transactions on Industrial Informatics*, 2017, 13(4): 2019–2029. doi: 10.1109/TII.2017.2685689.

[19] XIE Xin, WANG Heng, YU Lei, *et al*. Online algorithms for optimizing age of information in the IoT systems with multi-slot status delivery[J]. *IEEE Wireless Communications Letters*, 2021, 10(5): 971–975. doi: 10.1109/LWC.2021.3052569.

[20] LI Chengzhang, LI Shaoran, CHEN Yongce, *et al*. Minimizing age of information under general models for IoT

data collection[J]. *IEEE Transactions on Network Science and Engineering*, 2020, 7(4): 2256–2270. doi: 10.1109/TNSE.2019.2952764.

[21] KAM C, KOMPELLA S, NGUYEN G D, *et al*. On the age of information with packet deadlines[J]. *IEEE Transactions on Information Theory*, 2018, 64(9): 6419–6428. doi: 10.1109/TIT.2018.2818739.

[22] 王恒, 陈鹏飞, 王平. 面向WIA-PA工业无线传感器网络的确定性调度算法[J]. 电子学报, 2018, 46(1): 68–74. doi: 10.3969/j.issn.0372-2112.2018.01.010.

WANG Heng, CHEN Pengfei, and WANG Ping. Deterministic scheduling algorithms for WIA-PA industrial wireless sensor networks[J]. *Acta Electronica Sinica*, 2018, 46(1): 68–74. doi: 10.3969/j.issn.0372-2112.2018.01.010.

[23] BERTSEKAS D P. Dynamic Programming and Optimal Control[M]. 4th ed. Belmont: Athena Scientific, 2012.

[24] MNIH V, KAVUKCUOGLU K, SILVER D, *et al*. Human-level control through deep reinforcement learning[J]. *Nature*, 2015, 518(7540): 529–533. doi: 10.1038/nature14236.

[25] ABD-ELMAGID M A, DHILLON H S, and PAPPAS N. A reinforcement learning framework for optimizing age of information in RF-powered communication systems[J]. *IEEE Transactions on Communications*, 2020, 68(8): 4747–4760. doi: 10.1109/TCOMM.2020.2991992.

王　恒：男，教授，博士生导师，研究方向为工业物联网、时钟同步、实时调度等.

余　蕾：男，硕士生，研究方向为无线网络调度.

谢　鑫：男，博士生，研究方向为无线网络调度.

责任编辑：余　蓉