计算机系统应用 ISSN 1003-3254, CODEN CSAOBN
Computer Systems & Applications,2023,32(9):272−279 [doi: 10.15888/j.cnki.csa.009239]

E-mail: csa@iscas.ac.cn
http://www.c-s-a.org.cn
Tel: +86-10-62661041

基于 Q 学习的蚁群优化水声网络协议①

廖学文, 耿　烜

(上海海事大学 信息工程学院, 上海 200120)
通信作者: 廖学文, E-mail: 1464660968@qq.com

摘　要: 针对水声通信中数据传输延时高且动态适应性弱的问题, 提出了一种基于 Q 学习优化的蚁群智能水声网络路由协议 (Q-learning ant colony optimization, QACO). 协议包括路由行为和智能决策部分, 在路发现和维护阶段, 依靠网络智能蚂蚁进行网络拓扑环境的构建和节点之间的信息交换以及网络的维护. 在 Q 学习阶段, 通过定量化节点能量和深度以及网络传输延时学习特征作为折扣因子和学习率, 以延长网络的生命周期, 降低系统能耗和延时. 最后通过水声网络环境进行仿真, 实验结果表明 QACO 在能耗、延迟和网络生命周期方面都优于基于 Q 学习辅助的蚁群算法 (Q-learning aided ant colony routing protocol, QLACO) 和基于 Q-learning 的节能和生命周期感知路由协议 (Q-learning-based energy-efficient and lifetime-aware routing protocol, QELAR) 和基于深度路由协议 (depth-based routing, DBR) 算法.

关键词: 路由协议; Q 学习; 蚁群优化; 网络自适应; 多跳路由; 能耗优化; 遗传算法; 强化学习

引用格式: 廖学文,耿烜.基于 Q 学习的蚁群优化水声网络协议.计算机系统应用,2023,32(9):272–279. http://www.c-s-a.org.cn/1003-3254/9239.html

Ant Colony Optimization Based on Q-learning for Underwater Acoustic Network Protocol

LIAO Xue-Wen, GENG Xuan

(College of Information Engineering, Shanghai Maritime University, Shanghai 200120, China)

Abstract: To solve the problems such as high data transmission delay and weak dynamic adaptability of underwater acoustic communication, this study proposes an intelligent underwater acoustic network routing protocol based on Q-learning ant colony optimization (QACO). The protocol includes routing behavior and intelligent decision. In the route discovery and maintenance phase, the construction of the network topology environment and information exchange among nodes as well as the network maintenance are carried out by intelligent NetAnts. In the Q-learning phase, the node energy and depth and network transmission delay learning characteristics are quantified as discount factors and learning rates to extend the network lifecycle and reduce system energy consumption and delay. Finally, simulations are carried out through the underwater acoustic network environment, and the experimental results show that QACO outperforms the Q-learning aided ant colony routing protocol (QLACO), Q-learning-based energy-efficient and lifetime-aware routing protocol (QELAR), and depth-based routing (DBR) algorithm in terms of energy consumption, delay, and network lifecycle.

Key words: routing protocol; Q-learning; ant colony optimization (ACO); network adaptation; multi-hop routing; energy consumption optimization; genetic algorithm; reinforcement learning

① 基金项目: 上海市教委科技创新项目 (2101070010E00121)
收稿时间: 2023-03-03; 修改时间: 2023-04-04; 采用时间: 2023-04-20; csa 在线出版时间: 2023-07-21
CNKI 网络首发时间: 2023-07-24

随着我们对海洋探索的不断推进, 对水下传感网络 (underwater sensor networks, UWSNs) 的研究越来越受到研究者的青睐. 陆地通信与海洋通信有着本质的不同, 陆地通信所用的电磁波在水下传输具有严重的衰减, 水下环境更适用于声波进行传输. 并且水声通信在水中遭受着严重的频率相关衰减, 这就导致声学调制解调器的传输功率消耗要比陆地通信高得多[1−3]. 因此传统的陆地通信协议, 并不适用于水声通信. 此外, 在水下实现通信, 还会有声波的传输高延时; 水中节点会动态漂移造成链路断裂; 水下节点能量有限等问题.

为了克服 UWSNs 的问题, 已经提出了许多路由协议来解决在水下传感网络中数据传输的问题. 许多经典传统算法被提了出来, 如基于向量的转发协议[4]、定向洪泛基于路由协议[5] 和基于深度的协议 DBR[6] 等. 然而, 传统的优化方法受到在实践中难以获得的先验信道统计知识的限制, 很容易陷入局部最优和水下路由空洞问题[7,8], 因此, 转向智能自适应的在线学习方法非常重要.

为了解决水下路由局部最优和水下路由空洞问题, 研究员们开始转向对智能算法的研究, 例如蚁群优化 (ant colony optimization , ACO) 和 (Q-learning, QL) 强化学习算法[9]. ACO 是 Dorigo 等人在 1996 年引入的一种基于群体的元启发式方法[10]. 蚁群可以通过跟踪信息素来找到它们的巢穴和食物来源之间的最短路径. 然后, 较短路径上的蚂蚁比较长路径上的蚂蚁更早到达目的地. 此外, 强化学习算法 (reinforcement learning, RL)[11] 作为一种自适应算法, 可以通过学习训练解决水下路由局部最优和水下路由空洞问题, 实现水下更节能高效的路由机制. 目前, 强化学习已被证明是解决无线通信网络中各种问题的有效在线学习技术. 通过即时与外部环境交互的奖励反馈, RL 智能体可以在没有任何先验系统知识的情况下学习最优策略. 如基于 Q-learning 的节能和生命周期感知路由协议 (Q-learning-based energy-efficient and lifetime-aware routing protocol, QELAR)[12], 该协议根据节点的能量定义了其奖励函数. QELAR 虽然在能效方面表现良好, 但缺乏深度和延迟信息来优化路由程序, 使得节点容易陷入局部最优. 基于 Q 学习辅助的蚁群算法 (Q-learning aided ant colony routing protocol, QLACO)[13], 该协议根据节点能量、深度和节点连接稳定系数定义奖励函数, 并通过给单个节点添加等待时间的方式解决空洞问题, 但是由于数据传输经过节点的等待时间是不同的, 容易使得节点陷入局部最优, 降低网络性能.

因此, 本文基于蚁群优化 (ACO) 算法以及 QL 强化学习算法, 提出一种基于 Q 学习辅助的蚁群算法 (Q-learning ant colony optimization, QACO), 用以实现水下通信路由协议. 将节点深度、能量、传输延迟信息和水深链路质量 (link quality indicator, LQI)[14] 加入到奖励参数中, 使得奖励机制的评判更加全面. 并运用 ACO 实现信息的路由以及链路奖励的更新, 使得网络能在更短的时间找到最优路径, 提升网络的性能.

1 系统模型及问题表述

1.1 水下传感网络系统模型

水声网络传感器的水下部署见图 1, 分别有水底数据采集节点、水中移动 AUV, 以及水面数据接收节点构成. 假设水下网络一共有W个传感器节点, 在每个时隙中, 源节点将采集的数据从有限集$F=\{f_1,f_2,\cdots,f_h\}$发送数据到有限集为$P=\{p_1,p_2,\cdots,p_m\}$的中继节点, 数据包再经过中继节点被传递到水面目标节点$Z=\{z_1,z_2,\cdots,z_l\}$. F_h、P_m、Z_l分别表示为源节点、中继节点、目标节点. 每类节点对应的个数为$|F_h|\triangleq W_h$、$|P_m|\triangleq W_m$、$|Z_l|\triangleq W_l$, 即有$W=W_h+W_m+W_l$. 假设每个传感器节点的初始能量为E_{init}. 则$d_{i,j}$为空间节点$i(x_1,y_1,z_1)$和节点$j(x_2,y_2,z_2)$之间的距离为欧氏距离, 公式为:

$$d_{i,j}=\sqrt{(x_1-x_2)^2+(y_1-y_2)^2+(z_1-z_2)^2} \tag{1}$$

且每个节点的发射功率和接收功率分别为P_s和P_r, 当节点处于空闲状态时其功率为P_k, 每个节点都配置有声波调制解调器, 其最大可传输距离为R_0.

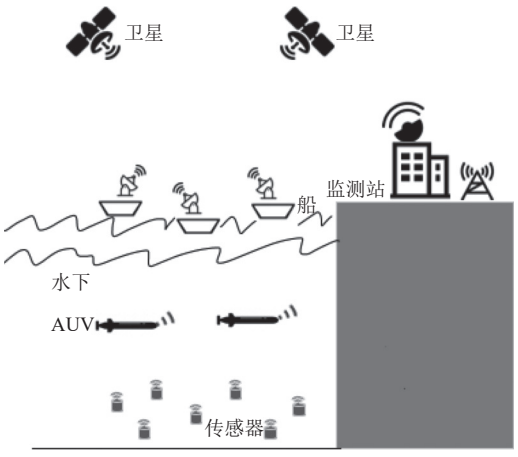

图 1 水声网络示意图

1.2 水下声学信道分析

不同于陆地通信, 水下是靠水声信道传输数据的. 因此, 水声信道的传输路径选择尤为重要. 不同的传输路径会导致不同的平均信噪比 (SNR)[15] 和链路吞吐量, 从而影响链路的能量消耗.

水声信道增益通常由距离和频率相关的大尺度衰落和时变多径衰落来描述. 信道增益A是传播距离d (m)的函数, 频率f (kHz), 以及一个随机变量γ (小规模衰落). 水声信道的路径传输损耗可表示为:

$$A(d,f)=\gamma\bar{A}(d,f) \tag{2}$$

其中, $\bar{A}(d,f)$称为平均信道功率增益. 根据 Deng 等人[16]的研究得:

$$\bar{A}(d,f)=d^k a(f)^d \tag{3}$$

其中, $a(f)$为吸收损耗系数, 单位为dB/km, 与频率有关. k为描述传播几何形状扩展因子. 根据Throp传输模型的经验公式[17], $a(f)$可以表示为:

$$10\lg a(f)=\frac{0.11f^2}{1+f^2}+\frac{44f^2}{4100+f^2}+2.75\times10^{-4}f^2+0.003 \tag{4}$$

$\lg a(f)$为以 10 为底的对数函数. 因此, 每次传输的瞬时信道增益随机变化, 从而影响瞬时接收信噪比和误码率 (BER). 假设节点在时隙t以频率f, 传输距离d, 发送一个数据包, 则成功传输的概率为:

$$\theta_{f,d}=\Pr\left\{BER_{f,d}\leqslant\delta\right\} \tag{5}$$

其中, δ为接收器可接受的最大误码率.

除了水下路径损耗之外, 环境噪声也会对节点的能量消耗造成影响. 水下环境噪声主要分为 4 种类型: 湍流噪声、船舶噪声、海浪噪声和热噪声. 假设 4 种噪声的功率谱密度和为: $N(f)$, 因此, 一个节点以频率f, 功率P, 发送数据, 声波传输距离为d的平均窄带 SNR 为:

$$SNR(d,f)=\frac{P/A(d,f)}{N(f)\Delta f} \tag{6}$$

其中, Δf为节点噪声带宽, 通常近似为一个常数. 根据通信理论可知, BER 会随着$SNR(d,f)$的增加而降低, 这样就有了传输距离d、频率f和传输成功率概率的关系:

$$P_{\text{success}}=1-\frac{1}{2}\left(1-\sqrt{\frac{SNR(d,f)}{1+SNR(d,f)}}\right) \tag{7}$$

1.3 水深链路质量

水声网络中, 有可能存在某一节点失效, 或者某一节点能量耗尽, 造成链路断裂的情况. 并引发路有空洞问题. 因此建立一个数学模型, 及时的评估链路质量, 以便尽早地发现问题, 提高系统的稳定性. 根据第 1.2 节定义的水下网络模型, 假设$M_{i,j}$为节点p_i和节点p_j之间的链路质量, 且$d_{i,j}<R_0$. 需要考虑节点间水声链路质量, 否则链路断裂, 质量指标为零, 无需考虑链路影响. 为了评估节点链路的优劣, 本文使用链路质量链路 LQI 越大, 则反映出该链路的质量越佳. 节点p_i和p_j的链路质量可以表示为:

$$M_{i,j}=\begin{cases}0, & d_{i,j}>R_0\\ a_1\times\left[\ln\left(R_0/\left(b_1^2\cdot d_{ij}\right)^2+b_1\right]\right., & 0<d_{i,j}\leqslant R_0\\ a_1\cdot b_1 & d_{i,j}=0\end{cases} \tag{8}$$

其中, $a_1\in N^*$, $b_1\in N^*$分别为链路调节系数.

1.4 问题描述

在 UWSNs 的数据传输过程中, 一条合适的传输路径受到传输节点深度、能量、延时的影响. 因此, 本文的设计的路由协议优化目标将围绕以上 3 点展开. 将路由协议基于蚁群算法的实现方式, 运用 QL 的决策设计理念. 将网络中央处理器看成一个集中式智能体, 在数据传输过程中, 不断地从环境中获取经验, 给出决策, 实现最优的动作. 最终实现一种路由, 以较低的数据传输延时均衡网络中的能量消耗. 假设N次传输数据所每次经过的节点组成的动作空间集合为ζ_e, 其中第n次传输经过了n_{ζ_e}个节点. 本节制定了一个评定指标V来评估在t时隙, 合集ζ_e中节点深度、能量和延时所对应的网络性能, 即:

$$V_n=\sum_{n=1}^{N}D_{(n,t)}/E_{(n,t)}T_{(n,t)} \tag{9}$$

其中,

$$D_{(n,t)}=D_z\left|\sum_{i=1,j=2}^{n_{\zeta e}}d_{ij}\right. \tag{10}$$

$$EX=\sum_{j=1}^{n_{\zeta e}}\left(E_{(j,t)}-E_{\text{init}}\right)/n_{\zeta_e} \tag{11}$$

$$E_{(n,t)}=\sum_{j=1}^{n_{\zeta e}}\left(\left(E_{\text{init}}-E_{(j,t)}\right)-EX\right)^2/n_{\zeta_e} \tag{12}$$

$$T_{(n,t)}=\left(\sum_{j=1}^{n_{\zeta e}}\left(1-\left(E_{(j,t)}/E_{\text{init}}\right)\right)\times T_{\max}+\eta\right)/n_{\zeta_e} \tag{13}$$

其中, $D_{(n,t)}$ 、$E_{(n,t)}$ 、$T_{(n,t)}$ 为第 n 次传输的节点深度、能量、延时的评定指标, D_z 为源节点的垂直深度, E_{init} 为节点初始能量, $T_{\max}$ 为最大持续时间, η 为随机值, EX 为节点平均能量和平均延时. 本算法研究就是找到一种合适的路由转发策略, 使得每条传输链路的 V_n 值最大, 表达式为:

$$\pi_* = \arg\max \sum_{n\in N} V_n \tag{14}$$

2 基于 QACO 的强化路由协议

QACO 的实现主要有两部分, 第 1 部分为做决策部分, 由 QL 强化学习算法完成; 第 2 部分为路由行为的实现, 由蚁群算法完成. 本节将对这两部分进行详细阐述.

2.1 强化学习及奖励机制

2.1.1 强化学习框架

RL 是机器学习的一个重要分支, 它为马尔可夫决策过程 (MDP) 提供了最优策略. 具体来说, 在 MDP 范式中, 智能体在时隙 t 观察环境状态 s_t, 然后在策略 π 下决定其动作 a_t. 受根据执行的 a_t 的影响. 环境将即时奖励 r_{t+1} 反馈给智能体, 并以转移概率 $P^{a_t}_{s_t s_{t+1}}$, 进入到新的状态 s_{t+1}, 其奖励 r_t 定义为:

$$r_t = P^{a_t}_{s_t s_{t+1}} r^{a_t}_{s_t s_{t+1}} \tag{15}$$

MDP 的目标是找到最优策略 π^*, 使得长期累积奖励最大化, 长期累积奖励定义为:

$$R_t = \sum_{k=t+1}^{\infty} \gamma^{k-t-1} r_k \tag{16}$$

其中, $\gamma \in [0,1]$ 为折扣因子. 由于智能体通常不知道先验概率的情况下, 就能够做出决策, 因此, QL 算法广泛用于解决概率未知的环境模型中. QL 使用 Q 表存储动作值函数 $Q(s,a)$. 在每个时隙 t 智能体根据状态 s_t 采取行动 a_t, 通过使用 ϵ-$greedy$ 策略[18], 获得奖励 r_{t+1}, 进入新状态 s_{t+1}, 其中对应的 $Q(s,a)$ 更新公式如下:

$$Q_\pi(s,a) = R_t + \gamma \sum_{s_{t+1}\in S} P^{a_t}_{s_t s_{t+1}} Q(s_{t+1}, a_t) \tag{17}$$

通过策略学习, Q 值最终收敛于最优 Q 函数, 定义为 $Q^*(s,a)$.

2.1.2 QACO 奖励机制

在水声环境中, 数据的传输包括成功和失败. 因此环境分别对成功和失败反馈奖励. 因此, 本文定义在某个时隙 t, 数据从节点 i 传输到节点 j 时, 奖励为:

$$R^*_{i,j} = P_{i,j} R_{i,j} + \widetilde{P}_{i,j}\, \widetilde{R}_{i,j} \tag{18}$$

其中, $P_{i,j}$ 、$R_{i,j}$ 分别为数据成功从节点 i 传输到节点 j, 对应的转移概率和奖励. $\tilde{P}_{i,j}$ 、$\tilde{R}_{i,j}$ 分别为数据节点 i 传输到节点 j 发送失败, 对应的转移概率和奖励. 由于数据只有成功和失败两种情况, 因此 $P_{i,j} = 1 - \widetilde{P}_{i,j}$, 由此得出环境转移概率为:

$$P^{a_t}_{s_t s_{t+1}} = \sum_{i=1,j=1,i\neq j}^{n_{\zeta_e}} P_{i,j} / n_{\zeta_e} \tag{19}$$

由式 (7) 得 $P_{i,j} = P_{\text{success}}$. 当信息包成功转发时, 在奖励函数中同时考虑了发送节点和接收节点的情况. 定义奖励为:

$$R_{i,j} = \alpha_1 E_{i,j} + \alpha_2 D + \alpha_3 T_{i,j} + M_{i,j} \tag{20}$$

其中, $E_{i,j}$ 为剩余能量, D 为节点深度, $T_{i,j}$ 为节点传输延时. $M_{i,j}$ 为节点之间链路质量. α_1, α_2, α_3 为各个奖励的权重系数. 其中 $E_{i,j}$ 定义为:

$$E_{i,j} = -\left(c_i + c_j\right) \tag{21}$$

其中,

$$c_i = 1 - \frac{E_i}{E_{\text{init}}} \tag{22}$$

式 (21) 中, c_i, c_j 分别为节点 i, j 的剩余能量价值, c_i 的剩余价值计算如式 (22) 所示, 同理可得 c_j. D 定义为:

$$D = -\frac{\left|D_{ix} - D_{jx}\right|}{D_{ij}} \tag{23}$$

其中, D_{ix}, D_{jx} 分别为节点 i, j 分别在 x 轴方向上的投影. D_{ij} 为节点 i, j 的欧氏距离. 节点传输延时为:

$$T_{ij} = T_i + T_j \tag{24}$$

$$T_i = (1 - (E_i / E_{\text{init}})) \times T_{\max} + \eta \tag{25}$$

其中, E_i 为当前节点能量, E_{init} 为节点初始能量, $T_{\max}$ 为最大持续时间, η 为随机值. 当传输失败时, 将不会再考虑节点 j 的能量作为奖励, 以及 j 点的延时信息, 因此节点传输失败奖励为:

$$\widetilde{R}_{i,j} = \alpha_1 c_i + \alpha_2 D + \alpha_3 T_i + M_{i,j} \tag{26}$$

综上所述, 第 n 次传输经过了 n_{ζ_e} 个节点, Q 网络的奖励为:

$$R_t = \sum_{i=1,j=1,i\neq j}^{n_{\zeta_e}} R^*_{i,j} \tag{27}$$

2.2 QACO 路由行为

QACO 通过将进行节点之间信息交互的数据包看作蚂蚁, 在网络中进行传播, 维护网络环境. 路由行为的实现, 主要分为 4 个阶段: 向前蚂蚁阶段、回溯蚂蚁阶段、路由维持阶段、数据传输阶段.

2.2.1 向前蚂蚁阶段

向前蚂蚁阶段: 发生在路由发现, 建立网络拓扑结构期间. 是一个全链路探测阶段, 蚂蚁由源节点出发, 采用ε-greedy算法不断地向前探查, 以找到通往目的节点的路径. 向前探索蚂蚁作为一种智能体去建立从源节点到目的节点的信息素轨迹. 蚂蚁经过空间节点i, j携带的节点信息为:

(1) 经过节点的剩余能量$E_{i,j}$.

(2) 蚂蚁经过节点的累积队列延迟$T_{i,j}$, 以及节点的深度信息D_i、D_j.

节点j能否作为节点i的转发节点由到达指标Λ决定.

$$\Lambda=\begin{cases}0, & d_{i,j}>R_0 \text{或} d_{i,j}=0\\ 1, & 0<d_{i,j}\leqslant R_0\end{cases} \tag{28}$$

$d_{i,j}$为节点i, j之间的欧氏距离, 若Λ为 1 则节点i将节点j加入邻居表, 否则不加入. 为了找到一条合适的路径, 蚂蚁需要带着数据包, 并不断地进行广播. 如图 2 为一个蚂蚁经过一个节点的过程图. 当一个蚂蚁经过一个节点, 首先判断达指标Λ, 若满足则将节纳入邻居节点, 继续向下传播, 否则直接丢弃. 若到达目的节点, 则是直接产生回溯蚂蚁, 更新链路信息素, 否则继续搜集信息向下转发.

2.2.2 回溯蚂蚁阶段

在t时隙向前蚂蚁携带数据信息到达目的节点时, Q 网络立即采用评估模型对第g条路径进行评估, 如式 (29) 所示:

$$\Psi_g=\hat{\Gamma}\times \mathrm{e}^{Q(s_t,a_t)+\gamma\sum\limits_{s_{t+1}\in S}P^{a_t}_{s_ts_{t+1}}Q(s_{t+1},a_t)} \tag{29}$$

其中,

$$\hat{\Gamma}=\sum\Lambda\Big/n_{\varsigma_e}(\Lambda=1) \tag{30}$$

其中, $Q(s_t,a_t)$为第g条路径的Q值奖励, γ为挥发系数, $P^{a_t}_{s_ts_{t+1}}$为转移概率, $\hat{\Gamma}$为平均邻居节点系数. 回溯蚂蚁通过评价模型对Ψ_g值进行评估. 最后选择最优的路径将数据发送相应数据包, 表达式为:

$$\psi_*=\arg\max\sum\Psi_n \tag{31}$$

其他向前蚂蚁变为回溯蚂蚁, 将路径的Q值信息带回链路中, 并更新链路的信息素.

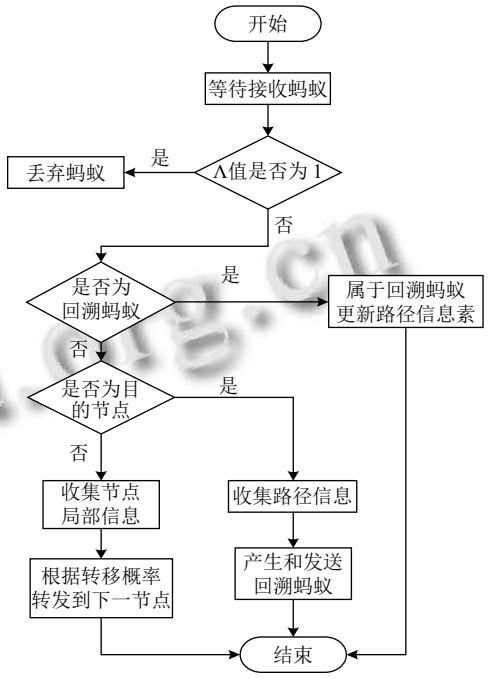

图 2 向前蚂蚁探索过程

2.2.3 路由维持和数据传输阶段

当源节点以一定的流量不断地发送数据时, 向前蚂蚁周期性地产生, 由于水下网络拓扑环境不是单一固定的, 是实时变化的, 因此在路由维持阶段, 还要处理网络拥堵问题. 网络拥堵情况由指标ϕ决定, 计算方式如下:

$$\phi=1-(N_{\text{receive}}/N_{\text{send}}) \tag{32}$$

其中, N_{receive}为接收节点个数, N_{send}为源节点发送节点数, ϕ值越大, 说明网络越空闲, 可以适当增加发送节点数, 否则应该降低节点数. 在算法 1 中概括了 QACO 算法, 数据的传输是选择式 (27) 最优决策进行数据的传输. 详细的传输方法及决策过程如算法 1 所示.

3 仿真实验与结果分析

3.1 仿真参数设置

为了评估 QACO 协议的性能, 本文构建了水下传感网络仿真环境, 并将传统 DBR 算法与本节中的算法相比较. 具体的仿真参数设置如下: 空间为一个3000×3000×3000 m^3的水下传输模型, 并设置 200–600 个传感器节点, 不规则地分布于环境模型中. 每个节点的最

大传输半径为 1 000 m; 在水面和底部分别设置一个源节点和目的节点. 维护蚂蚁周期性的广播数据包以维护传感网络, 网络仿真参数如表 1 所示. $\alpha_1,\alpha_2,\alpha_3$分别为节点能量、深度、延时的调整系数, α_1水深系统设计的调整参数, 用于将数据从源节点传到目的节点.

算法 1. QACO 路由决策算法

输入: 初始化 Q-learning 参数: 折扣因子γ, 奖励值R, 选择概率P, 权重系数$\alpha_1,\alpha_2,\alpha_3$.

输出: 式 (14) 得出的路由决策.

● 初始化阶段

1. 初始化节点参数: 中继节点W_m, 节点传输范围R_0, 节点初始能量E_{init}, 网络节点发射功率P_s和接收功率P_r, 以及空闲时的功率P_k.

2. 初始化 Q 网络相关参数: 折扣因子γ, 奖励值R, 选择概率P, 权重系数$\alpha_1,\alpha_2,\alpha_3$, 以及初始$Q$矩阵, 数据包转发次数$G_0$.

● QACO 路由及决策

3. for $episode$=1,2,3,$\cdots$,G_0 do//设置数据传输次数

4. for t=1,2,3,$\cdots$,T do //传输时隙

5. 向前蚂蚁依次从源节点出发, 广播节点W_n,W_l包头信息.

6. 基于ε-$greedy$访问节点, 并将已访问节点加入自己的禁忌表.

7. 计算相邻i,j节点之间的欧氏距离d_{ij}.

8. 若$0<d_{ij}<R_0$, 则将节点分别纳入自己的邻居表. 否则将不加入邻居表.

9. 计算第n次传输的环境状态$s_{(n,t)}=[D_{(n,t)},E_{(n,t)},T_{(n,t)}]$, $n\in G_0$.

10. 将状态输入 Q 网络计算 Q 值

11. 根据评估模型对第传到目的节点的g条路径进行评估得出ψ_*.

12. 最优路径回溯蚂蚁向源节点传回响应信息, 其余回溯蚂蚁则分别返回更新链路信息.

13. 信息更新完成, 选择动作a_t选出的最大 Q 值的路径进行下一次数据的发送.

14. 以式 (27) 更新计算动作的奖励R_{ij}^*, 且状态转变为s_{t+1}.

15. end for

16. end for

表 1 系统仿真参数

序号	主要参数	量值
1	节点数 (W)	200–600
2	三维空间 (V)	$3000\times3000\times3000$ m^3
3	初始能量 (E_{init})	6 000 J
4	发射功率 (P_s)	10 W
5	接收功率 (P_r)	3 W
6	空闲功率 (P_k)	20 mW
7	声信号频率 (f)	25 kHz
8	传输范围 (R_0)	1 000 m
9	模拟时间 (T_{sim})	500 s
10	最大迭代次数 (次)	2 000

3.2 仿真结果分析

如图 3 所示, 首先对不同α_1下的算法的平均能耗进行实验, 平均能耗进行归一化为(0,1). 在进行模拟实验中$\alpha_1\in[0,0.6]$、$\alpha_2=\{0,0.1,0.2\}$、$\alpha_3=\{0,0.1,0.2\}$. 通过对比不同的参数进行仿真实验, 可以得出结论: α_1越高, 并且α_3越低, 会使得网络的平均能耗越低, 网络的寿命越长.

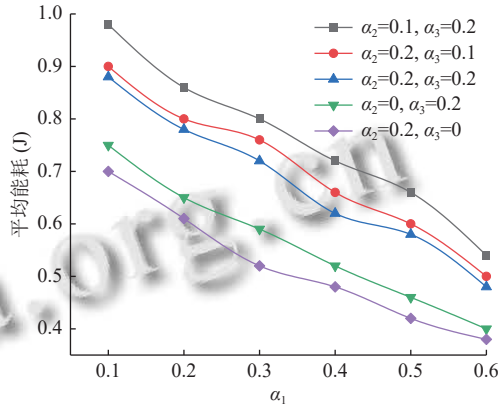

图 3 不同α_1下的平均能耗

由图 3 可知, 得出了不同权重下对网络能耗的影响情况, 为了使得网络性能更加, 取 $\alpha_1=\alpha_2=\alpha_3=0.2$ 作为 QACO 的权重系数, 分别与其他同类算法相互比较. 由图 4 显示了不同节点密度情况下, 网络中节点平均延时情况. 由于 QACO 引入了水深链路质量因子 LQI, 相比于 QLACO 单个节点添加等待时间的方式解决空洞的方式, QACO 带来了更低的节点延时信息. 并且随着节点的增加, QACO 的优势显得更加明显. 同时与考虑单一影响因素的协议相比, QACO 协议更稳定且延时更低.

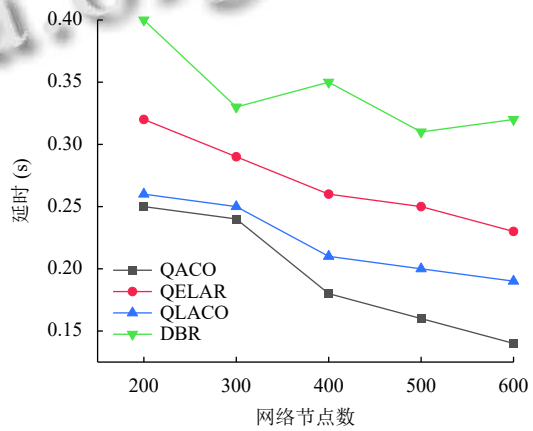

图 4 网络节点平均延时

结合图 5 和图 6, 随着网络节点数量的增加, 采用 QACO 路由协议和与 QELAR、QLACO 和 DBR 协议相比, 具有更长的网络生命周期和最少的能耗. 由于

QLACO 在处理路由空洞时会造成节点长时间的等待, 增加节点能量的消耗, 容易使节点陷入局部最优, 并且在节点数量增加时这一弊端显得更为明显. 而 QACO 通过链路质量因子 LQI 来判断链路好坏, 并兼顾考虑节点深度、能量、延时, 使得路由得出全局最优选择, 避免了路有空洞问题. 同时 QACO 采用智能蚂蚁进行路由行为, 优化了链路的传输效率, 降低了系统能耗, 延长了网络的生命周期.

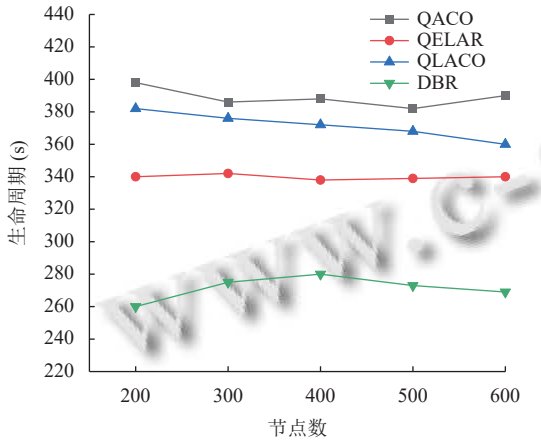

图 5 网络生命周期图

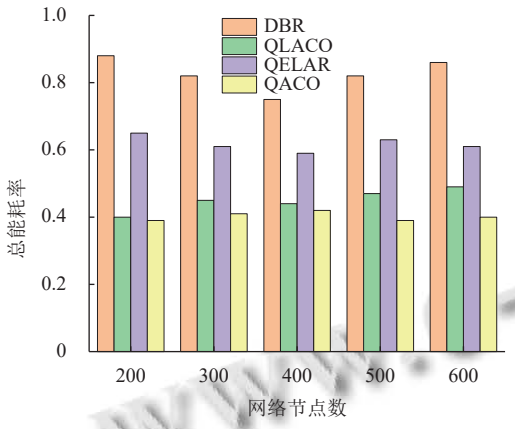

图 6 网络的能耗比

4 结论

本文提出了一种基于 Q 学习优化的蚁群智能水声网络路由协议, 利用 Q 学习进行确定全局路由最优选择, 并设计基于延迟、剩余能量、深度和链路质量相关奖励函数. 并将蚁群算法应用于水声传感网络的路由行为, 实现水下数据的稳定可靠传输, 并解决了路有空洞问题, 分别从网络的延时、能量消耗、生命周期方面对协议进行了仿真实验, 实验证明, QACO 算法在延时, 能耗以及网络的生命周期方面都明显优于 QELAR、QLACO 和 DBR 算法.

参考文献

1 Zhang RX, Ma XL, Wang DQ, *et al*. Adaptive coding and bit-power loading algorithms for underwater acoustic transmissions. IEEE Transactions on Wireless Communications, 2021, 20(9): 5798–5811. [doi: 10.1109/TWC.2021.3070363]

2 Doosti-Aref A, Ebrahimzadeh A. Adaptive relay selection and power allocation for OFDM cooperative underwater acoustic systems. IEEE Transactions on Mobile Computing, 2018, 17(1): 1–15.

3 Chen XK, Yang HJ, Wang QM, *et al*. Network coding-based relay selection and power distribution for underwater cooperative communication networks. Proceedings of the 2020 International Conference on Wireless Communications and Signal Processing. Nanjing: IEEE, 2020. 596–601.

4 Xie P, Cui HJ, Lao L. VBF: Vector-based forwarding protocol for underwater sensor networks. Proceedings of the 5th International Conference on Research in Networking. Coimbra: Springer, 2006. 1216–1221.

5 Hwang D, Kim D. DFR: Directional flooding-based routing protocol for underwater sensor networks. Proceedings of the OCEANS 2008. Quebec City: IEEE, 2008. 1–7.

6 Yan H, Shi ZJ, Cui JH. DBR: Depth-based routing for underwater sensor networks. Proceedings of the 7th International Conference on Research in Networking. Singapore: Springer, 2008. 72–86.

7 Yildiz HU, Gungor VC, Tavli B. Packet size optimization for lifetime maximization in underwater acoustic sensor networks. IEEE Transactions on Industrial Informatics, 2019, 15(2): 719–729. [doi: 10.1109/TII.2018.2841830]

8 Nazareth P, Chandavarkar BR. Link and void aware routing protocol for underwater acoustic sensor networks. Proceedings of the 12th International Conference on Computing Communication and Networking Technologies. Kharagpur: IEEE, 2021. 1–6.

9 Shi YM, Rong ZH. Analysis of Q-learning like algorithms through evolutionary game dynamics. IEEE Transactions on Circuits and Systems II: Express Briefs, 2022, 69(5): 2463–2467. [doi: 10.1109/TCSII.2022.3161655]

10 Dorigo M, Maniezzo V, Colorni A. Ant system optimization by a colony of cooperating agents. IEEE Transactions on Systems, Man, and Cybernetics, Part B (Cybernetics). 1996,

26(1): 29–41.

11 Lyu L, Shen Y, Zhang SC. The advance of reinforcement learning and deep reinforcement learning. Proceedings of the 2022 IEEE International Conference on Electrical Engineering, Big Data and Algorithms. Changchun: IEEE, 2022. 644–648.

12 Hu TS, Fei YS. QELAR: A Q-learning-based energy-efficient and lifetime-aware routing protocol for underwater sensor networks. Proceedings of the 2008 IEEE International Performance, Computing and Communications Conference. Austin: IEEE, 2008. 247–255.

13 Fang ZR, Wang JJ, Jiang CX, *et al*. QLACO: Q-learning aided ant colony routing protocol for underwater acoustic sensor networks. Proceedings of the 2020 IEEE Wireless Communications and Networking Conference. Seoul: IEEE, 2020. 1–6.

14 Malajner M, Gleich D. Study of link quality indicator for possible usage in angle of arrival estimation. Proceedings of the 21 International Conference on Systems, Signals and Image Processing. Dubrovnik: IEEE, 2014. 195–198.

15 Haneef N, Chandavarkar BR. Dependency of routing protocol on pH and SNR in underwater communications. Proceedings of the 12th International Conference on Computing Communication and Networking Technologies. Kharagpur: IEEE, 2021. 1–7.

16 Deng ZC, Yu X, Zhu ZB, *et al*. Adaptive kalman filter based single beacon underwater tracking with unknown effective sound velocity. Proceedings of the 8th IEEE International Conference on Underwater System Technology: Theory and Applications. Wuhan: IEEE, 2018. 1–5.

17 Brekhovskikh LM, Lysanov YP. Fundamentals of Ocean Acoustics. 3rd ed., New York: Springer-Verlag, 2003.

18 Bhandarkar AB, Jayaweera SK. Optimal trajectory learning for UAV-mounted mobile base stations using RL and greedy algorithms. Proceedings of the 17th International Conference on Wireless and Mobile Computing, Networking and Communications. Bologna: IEEE, 2021. 13–18.

(校对责编: 孙君艳)