

面向移动汇聚节点的无线传感器网络 自适应路由方法

杨 军^{1,2}, 秦锋蕴^{1,2}, 韩 晨^{1,2}

(1. 南瑞集团有限公司(国网电力科学研究院有限公司), 南京 211106;

2. 北京科东电力控制系统有限责任公司, 北京 100192)

摘要: 针对无线传感器网络面向移动汇聚节点的自适应路由问题, 为实现路由过程中对节点能量以及计算、存储、通信资源的优化利用, 并对数据传输时延和投递率等服务质量进行优化, 提出一种基于强化学习的自适应路由方法, 设计综合的奖赏函数以实现能量、时延和投递率等多个指标的综合优化; 从报文结构、路由初始化、路径选择等方面对路由协议进行详细设计, 采用汇聚节点声明以及周期性洪泛机制加速收敛速度, 从而支持汇聚节点的快速移动; 理论分析表明基于强化学习的路由方法具备收敛快、协议开销低以及存储计算需求小等特点, 能够适用于能量和资源受限的传感器节点; 在仿真平台中通过性能评估和对比分析验证了所述自适应路由算法的可行性和优越性。

关键词: 无线传感器网络; 汇聚节点; 自适应路由; 强化学习

Adaptive Routing for Wireless Sensor Network with a Mobile Sink Using Reinforcement Learning

Yang Jun^{1,2}, Qin Fengyun^{1,2}, Han Chen^{1,2}

(1. NARI Group Corporation (State Grid Electric Power Research Institute), Nanjing 211106, China;

2. Beijing Kedong Power Control System Co., Ltd., Beijing 100192, China)

Abstract: Many applications of wireless sensor network use a mobile sink to collect data detected by network nodes. To make routing approach in such scenarios better support sink mobility and be adaptive to varying network conditions, an adaptive routing approach based on reinforcement learning was proposed. A comprehensive reward function concerning mean hop count, residual energy and link quality was designed for learning to achieve a trade-off among multiple metrics of the network performance. Sink announcement was used to guarantee the convergence speed for obtaining an optimal route. Theoretical analysis shows that the approach is feasible for resources constrained sensor nodes due to its low overhead, small storage and computation requirements. Simulation results show the feasibility and superiority of the proposed approach.

Keywords: wireless sensor networks; mobile sink; adaptive routing; reinforcement learning

0 引言

无线传感器网络的路由协议负责确定源节点和目的节点之间的优化路径以及沿优化路径正确转发数据^[1]。面向固定汇聚节点的路由方法已经得到了广泛的研究, 由于能耗均衡的需求和移动设施的普及, 在大量应用中需要采用移动汇聚节点对监测区域的数据进行采集^[2-4], 所设计的路由机制需要适应由于汇聚节点移动带来的网络状态的不断变化^[2]。传感器节点的能量限制要求路由方法考虑节约能量和均衡能耗以延长网络的生存时间; 由于传感器节点具备有限的资源和通信带宽, 路由算法应该简单高效, 占用尽可能少的计算、存储和通信资源; 路由过程同时需要对数据传输的实时性和投递率等服务质量进行优化^[5]。

近年来, 国内外研究者对移动汇聚节点的研究进行了很大关注^[2]。文献 [6] 研究了双层数据发布 (TTDD) 协议, 通过将无线传感器网络分割为多个单元实现向多个移动汇聚节点的数据传输; 然而, 当网络数据传输率增大时, TTDD 的开销

会变得非常大, 因此其更适用于低负载和网络流量的场景。文献 [7] 提出了基于分簇的路由方法, 簇头节点聚合时空关联数据并将其转发至合适的终端节点, 终端节点作为代理节点位于移动汇聚节点的轨迹附近; 然而, 基于代理的方法会带来很大的时延, 且成本较高, 对分簇的假定也不适用于无线传感器网络的一般应用场景。文献 [8] 提出了一种基于地理位置的服务和路由方法。每当链路状况发生改变时, 汇聚节点的邻居节点都会更新位置信息, 该路由方法假定所有传感器节点都具备定位单元, 频繁的位置更新也增加了路由的开销。

上述路由方法自适应性不足, 为使路由方法适应快速变化的网络状况, 很多研究集中在了强化学习技术的应用^[9]。为提升网络状态变化时的路由效率, Boyan 和 Littman 首次将强化学习技术应用到数据包的路由问题中, 并提出了一种 Q-路由方法^[10]。由于强化学习具有简单性、自适应性和鲁棒性等特征^[11], 非常适用于无线传感器网络的路由问题^[12]。文献 [13] 基于强化学习方法提出了一种多播路由框架以高效的处理具有多个移动汇聚节点的无线传感器网络路由过程中的多播和移动性问题。上述方法适用于汇聚节点低速到中速的移动场景。文献 [14] 基于强化学习技术提出了一种适用于移动 AD Hoc 网络的主动路由协议, 该方法可以最大化节点生存时间和

收稿日期: 2017-12-20; 修回日期: 2018-01-18。

作者简介: 杨 军(1985-), 男, 河北衡水人, 博士, 主要从事测控技术与自动化系统方向的研究。

快速适应由于节点移动带来的网络拓扑变化；该方法针对移动 AD Hoc 网络并且假定所有的节点都具备 GPS 设备，不适用于无线传感器网络的一般应用场景。

可见，强化学习是处理无线传感器网络路由问题中节点移动性和自适应性的有效方法。针对当前方案不具备通用性和不能充分支持无线传感器网络高度自适应性的不足，需要综合考虑时延、投递率和网络生存时间等多个指标以获取最优的网络性能。本文提出一种基于强化学习的支持移动汇聚节点的自适应路由方法（learning based routing supporting mobile sink, LRMS）。建立路由问题基于强化学习的模型，并设计综合考虑跳数、链路质量和能量分布的奖赏函数对无线传感器网络的时延、投递率以及生存时间进行优化。理论和仿真评估验证了 LRMS 在无线传感器网络面向移动汇聚节点自适应路由问题中的可行性和优越性。

1 路由模型建立

Q 学习（Q-Learning, QL）^[15] 由 Watkins 在 1989 年提出，是马尔科夫环境下的模型无关的强化学习方法，对传感器节点的计算和存储资源有着中等的要求^[11]，可用于解决无线传感器网络的路由问题^[12]。在给定策略 π 下，Q 学习的每个状态/行动 (s_i, a_i) 都会对应一个 Q 值函数 $Q_\pi(s_i, a_i)$ 用于表示在状态 s_i 下采取行动 a_i 的累计奖赏情况。Q 值用于评价特定状态下采取某个行动的优劣。

在 Q 学习中，agent 可以通过不断试错和有延迟的奖赏获得最优的行动策略。Q 学习算法的核心是通过迭代更新 Q 值以获取最优的行动策略，更新过程定义如下式所示。

$$Q(s_t, a_t) \leftarrow Q(s_t, a_t)(1 - \alpha) + \alpha[r(a_t) + \gamma \max_{a_{t+1}} Q(s_{t+1}, a_{t+1})] \quad (1)$$

其中： $Q(s_t, a_t)$ 为在状态 s_t 下采取行动 a_t 所对应的 Q 值。 $r(a_t)$ 为即时的奖赏。 $\max_{a_{t+1}} Q(s_{t+1}, a_{t+1})$ 为当 agent 在下一个状态 s_{t+1} 下的最大可能 Q 值。 α 为学习率，学习率决定新获得的信息覆盖旧的信息的程度，较大的学习率意味着 Q 值更有可能按照新获得的奖赏进行较快的更新。 γ 为折扣因子，决定未来奖赏的重要性。

1.1 基于 Q 学习的路由模型

假定无线传感器网络由 p 个传感器节点和一个移动汇聚节点组成。在每个节点内实现基于 Q 学习的路由算法，将节点将要发送或转发的数据包作为 agent，通过 Q 表的查询选择具备最大 Q 值的路径将数据传输到汇聚节点。路由模型中 Q 学习的状态、行动、Q 值以及 Q 值的更新过程分别定义如下。

(1) 状态： $V = \{v_1, v_2, \dots, v_p\}$ 为无线传感器网络节点的集合， $S = \{s_1, s_2, \dots, s_p\}$ 为 agent 状态的集合。当数据包在第 i 个节点 v_i 时，相关的数据包状态为 s_i 。

(2) 行动： $V_i = \{v_1, v_2, \dots, v_k\}$ 为传感器节点 v_i 的邻居节点集合，各邻居节点对应的状态集合为 $S_i = \{s_1, s_2, \dots, s_k\}$ 。则定义 agent 在状态 s_i 下可采取的行动集合为 $A_i = \{a_1, a_2, \dots, a_k\}$ ，其中 $a_j \in A_i$ 表示将数据包从 s_i 转发至邻居节点 s_j ，当采取行动 a_j 之后，agent 状态转为 s_j 。

(3) Q 值：Q 值代表在各状态下采取每个行动的优劣，当数据包状态为 s_i 时，行动 a_j 对应的 Q 值为 $Q(s_i, a_j)$ ，表示通过采取行动 a_j 将数据包从 s_i 转发至 s_j 的质量。

(4) Q 值更新：为避免学习率过低带来的延迟，将学习率

α 设置为 1 以加速学习过程。当从 s_j 接收到奖赏信息后，状态/行动 (s_i, a_j) 的 Q 值将按以下公式进行更新。

$$Q(s_i, a_j) \leftarrow r(a_j) + \gamma \max_{a_k} Q(s_j, a_k) \quad (2)$$

其中： $r(a_i)$ 为奖赏函数， γ 为折扣因子，当 γ 较小时，agent 更加注重即时的奖赏；当 γ 较大时，agent 更加注重未来的奖赏。 $\max_{a_k} Q(s_j, a_k)$ 为在状态 s_j 下采取最优行动 a_k 对应的最大 Q 值。

1.2 奖赏函数设计

奖赏函数需要考虑时延、投递率以及能耗均衡等优化目标以获取最优的整体网络性能。时延主要由数据包从源节点到汇聚节点经过的跳数决定，能耗均衡可以由各节点剩余能量的分布决定，数据传输的投递率则直接受链路质量的影响，链路质量可以通过接收信号强度（RSS）计算得到。因此，奖赏函数需要考虑剩余跳数、剩余能量以及 RSS 三个指标。

当数据包通过采取行动 a_j 从 s_i 传递到 s_j 时，环境对状态/行动 (s_i, a_j) 的奖赏 $r(a_j)$ 为 s_j 对 s_i 的即时反馈。奖赏函数设计如下：

$$r(a_j) = \omega_1 \cdot H(s_j) + \omega_2 \cdot E(s_j) + \omega_3 \cdot L(s_j) \quad (3)$$

其中： $H(s_j)$ 、 $E(s_j)$ 和 $L(s_j)$ 分别为剩余跳数、剩余能量和 RSS 的归一化值， ω_1 、 ω_2 和 ω_3 为对应的权重。 $H(s_j)$ 、 $E(s_j)$ 和 $L(s_j)$ 的定义如下：

$$\begin{cases} H(s_j) = \sqrt{\frac{H_{rem}(s_j)}{H_{max}(s_j)}} \\ E(s_j) = \frac{E_{res}(s_j)}{E_{init}(s_j)} \\ L(s_j) = \frac{(|P_{min}| - |P_{Rr}(s_j)|)}{|P_{min}|} \end{cases} \quad (4)$$

$H_{rem}(s_j)$ 为节点 s_j 中记录的距离汇聚节点的剩余跳数， $H_{max}(s_j)$ 为无线传感器网络中的最大可能跳数。均方根可避免 $H(s_j)$ 取值过小以拉低奖赏函数的总值。 $E_{res}(s_j)$ 和 $E_{init}(s_j)$ 分别为节点 s_j 的剩余能量和初始能量，所有节点具有相同的初始能量 $E_{init}(s_j)$ 。 $P_{Rr}(s_j)$ 为 s_j 的 RSS，单位为 dBm。由于无线传感器网络的通信协议为短距离通信协议，其传输功率较低，同时由于路径损耗的存在， $P_{Rr}(s_j)$ 的值为负数。因此，RSS 的绝对值越小，说明信号强度越大，链路质量越好。 P_{min} 为 RSS 可能取得的最小值，单位为 dBm。此处，将传感器节点的接收灵敏度设置为 -100 dBm，可知 RSS 的实际取值必大于 -100。将 P_{min} 设置为 -200 dBm 以避免 $L(s_j)$ 的取值过小。

2 基于强化学习的路由方法设计

2.1 报文结构设计

路由信息更新及数据传输都有赖于携带用户数据以及奖赏信息的报文交换。通用报文结构如图 1 所述，包括包头和载荷。包头可分为 3 个区域：报文信息域，奖赏信息域，路由信息域。载荷区域为可选区域，包含报文所携带的应用层用户数据，通常为传感器节点采集的参数或控制命令。

2.2 路由表初始化

为满足应用的实时性，传感器节点到移动汇聚节点的初始路径必须在很短的时间内建立。然而，当直接应用无模型的 Q 学习时，由于需要对每个状态进行频繁的随机搜索，Q 值收敛速度非常慢，会造成路由由建立初始阶段的较大时延，不能满足应用的实时性需求。因此，此处采用汇聚节点声明为每个状态/行动建立初始 Q 值从而快速建立初始路径。

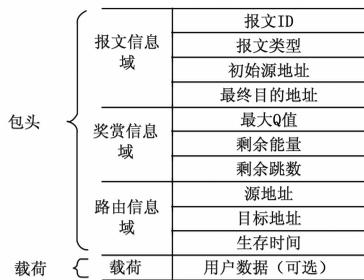


图 1 报文结构

在初始阶段, 传感器节点中存储的所有 Q 值均初始化为 0, 汇聚节点的最大 Q 值初始化为 1。当汇聚节点进入无线传感器网络范围, 会立即发送广播到整个网络的控制包作为节点声明过程。经过初始化和汇聚节点声明, 各节点会建立到汇聚节点的初始路径。

2.3 选择最优路径

传感器节点存储一个 Q 表以记录对应每个邻居节点的 Q 值, 下一跳节点可以通过查表操作得到。如图 2 示, $\{s_1, s_2, \dots, s_k\}$ 为节点 s_i 的邻居节点的集合, $\{a_1, a_2, \dots, a_k\}$ 为节点 s_i 可以采取的行动的集合, 每个行动对应一个邻居节点。对应于 s_j 的 Q 值是 $Q(s_i, a_j)$ 。当算法收敛, 具备最大 Q 值的行动将会被选择为最优的行动。由于传感器节点对每个邻居节点都保持一个 Q 值, 当网络状况发生变化时, 可以在不同的路径之间实现快速转换。

由于传感器节点对正确接收到的数据包进行确认, 确认包的头部包含赏赏信息用以更新源节点中对应的 Q 值。在数据传输过程中, 参与发送和转发数据包的节点都会实时更新其 Q 表以及路由信息。数据传输过程中 Q 表的动态更新可以进一步保证 LRMS 的自适应性和处理不断变化的网络环境。

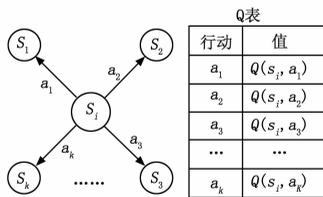


图 2 据 Q 表选择最优路径

2.4 移动性管理

通过不断学习网络环境的变化, 可从本质上支持汇聚节点的移动性, Q 值会在学习过程中根据赏赏函数进行不断更新。然而, 由于汇聚节点的高速移动, 同时由于学习延迟的存在, 在学习到最新路径之前, 当前路径可能会很快变得不再适用。因此, 仅仅通过在线学习和随机探索不能处理由于汇聚节点高速移动带来的网络结构快速变化。

为保证传感器节点路由信息的快速更新, 在汇聚节点中采用周期性的洪泛机制。移动汇聚节点以间隔 T_{Flood} 发送广播到整个网络的控制包。由于在控制包中含有赏赏信息, 每次洪泛过程中, 所有节点都将更新其关于汇聚节点的路由信息。这样, 每个节点的路由表会根据最新的汇聚节点信息实现更新。在动态的网络中需要更短的洪泛周期 T_{Flood} , 然而, 频繁的洪泛广播会消耗更多的能量和产生更大的开销。因此, 需要设置合理的 T_{Flood} 值以得到网络开销和反应速度之间平衡。

除周期性洪泛, 由于路径损耗 RSS 与各节点之间距离直接相关^[16], 当汇聚节点距离较远时, RSS 将会变小, 对应于汇聚节点的 Q 值会随着距离的增大而变小。当汇聚节点靠近时, 对应于汇聚节点的 Q 值会随着距离的减小而增大。因此, 汇聚节点的移动性可以通过 RSS 检测得到, 并通过 Q 值的更新使数据包更加倾向于通过距离汇聚节点较近的节点传送, 从而进一步保证汇聚节点移动时数据包的可靠传输。

3 路由算法评估验证

3.1 算法理论分析

3.1.1 收敛性分析

Watkins 和 Dayan 对 Q 学习的收敛性有如下证明: 只要每个状态下的所有行动都被采样, 而且每个行动/值都为离散表示, Q 学习就会以概率 1 收敛^[17]。假定无线传感器网络有 p 个传感器节点和 1 个移动汇聚节点, 任一传感器节点的单跳邻居节点最多有 k 个, 为使路由算法收敛, 所有传感器节点要采取的行动数量最多为 $(p+1) \times k$ 次。本算法中, 汇聚节点声明和周期性洪泛过程采取的行动总数为 $(p+1) \times k$, 且所有可能行动都被采样, 可保证路由收敛。

3.1.2 开销分析

根据报文结构, 单个确认包、控制包和数据包的开销均为 20 字节。假定在周期 T 内无线传感器网络传输的数据包数量为 S , 则数据包和确认包的总开销为 $2 \times S \times 20$ 字节。汇聚节点洪泛周期为 T_{Flood} , 由于每个节点仅对控制包执行单次接收和广播, 对于具有 p 个传感器节点和 1 个汇聚节点的网络, 假定每个传感器节点的邻居节点数量最多为 k 个。则在 T_{Flood} 内控制包的数量最多为 $(p+1) \times k$ 个。因此, 在周期 T 内控制包的开销为 $(\frac{T}{T_{Flood}}) \times (p+1) \times k \times 20$ 字节。综上, 在周期 T 内的

总开销为 $[2S + (\frac{T}{T_{Flood}}) \times (p+1) \times k] \times 20$ 字节。

3.1.3 存储和计算需求

每个传感器节点需建立一个 Q 表用以存储所有可能的行动及其对应的 Q 值, 每个行动和值分别用 16 比特的字来表示。对于有 k 个邻居的传感器节点来说, Q 表所需的存储量为 $32 \times k$ 比特。赏赏信息存储需求为 48 比特。因此, 传感器节点总存储需求为 $(32 \times k + 48)$ 比特。传感器节点需要执行计算操作以选择路径和更新 Q 值, 选择路径所需的计算量是 $O(k)$, 更新 Q 值所需的基本运算操作数量为常数, 其计算需求为 $O(1)$ 。综上, LRMS 中选择路径和更新 Q 值的总计算需求为 $O(k)$ 。

根据以上分析, 路由算法需要极小的存储和计算资源, 适用于存储和计算资源受限的无线传感器节点。

3.2 仿真环境设计

采用 OMNeT++ 仿真环境评估路由性能。传感器节点以复合模块的形式实现, 在应用层中周期性的产生数据包以模拟传感器节点的信息采集; 在网络层采用本文所述基于 Q 学习的路由算法为数据包选择优化路径; MAC 层和物理层采用符合 IEEE 802.15.4 标准的协议规范; 在物理层加入能耗模型以模拟射频模块不同动作的能量消耗。

无线传感器网络的仿真参数如表 1 所示。应用层中以 3 s 为周期生成常规数据包, 默认的学习率和折扣因子分别设置为 1.0 和 0.5; 汇聚节点的洪泛周期为 1 s; 赏赏函数中剩余跳

数、剩余能量以及链路质量的权重分别设置为 -1.0、0.1 和 0.3；默认传感器节点数量为 16 个；默认的汇聚节点移动速度为 20 m/s；单次运行的仿真持续时间为 1 000 s。

表 1 无线传感器网络的仿真参数设置

参数	值
数据传输速率	250 kbps
载波频率	2.4 GHz
路径损耗系数	2.5
传输距离	150 m
工作电压	3.3 V
初始能量	2 kJ
接收电流	18.8 mA
监听电流	18.8 mA
发送电流	27.4 mA

采用平均跳数、能量分布以及投递率 3 个指标评估算法性能。平均跳数用于评估数据包的时延，投递率用于评估数据传输的可靠性，能量分布用各节点平均功耗的标准差表示，用于评估网络生存时间。

3.3 算法性能评估

将 3 个指标的仿真结果均归一化到 [0, 1] 范围之内。平均跳数和能量分布的归一化值由各仿真结果分别除以其最大值得到，由于投递率本身就在 [0, 1] 范围之内，可以直接作为归一化值使用。

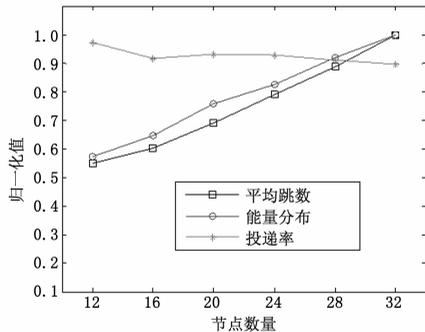


图 3 不同网络规模下的性能

图 3 为传感器节点数量以步长 4 在 12~32 之间变化时的性能。可见，平均跳数和能量分布随着网络规模的增大而成比例的增大，这是由于源节点到汇聚节点的路径长度会随着网络规模的增大而变长；投递率的减小并不明显。可见，LRMS 具有很好的可扩展性，并且在不同的网络规模下均取得了较好的性能。

图 4 分析了汇聚节点速度以步长 5 m/s 在 5~30 m/s 之间变化时对网络性能的影响。可见，随着汇聚节点速度的增加，平均跳数和能量分布的改变不明显，投递率有轻微减小；即使速度达到 30 m/s 时仍然保持了较好的性能。可见，LRMS 可适应较高的汇聚节点速度。

3.4 算法对比分析

为验证算法的优越性，在不同的网络规模和汇聚节点速度下对无线传感器网络的平均跳数、能量分布和投递率进行对比。在 OMNeT++ 中实现树状路由（tree based routing, TBR）^[18] 以进行对比分析，两种路由协议采用相同的网络参数以及初始化过程以保证对比的公平性。

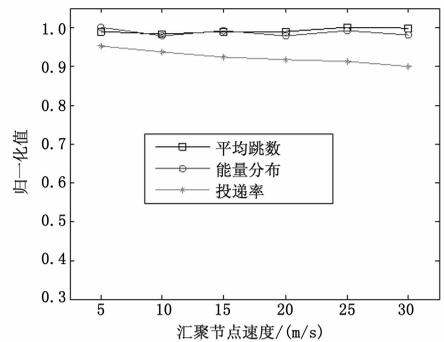


图 4 不同汇聚节点速度下的性能

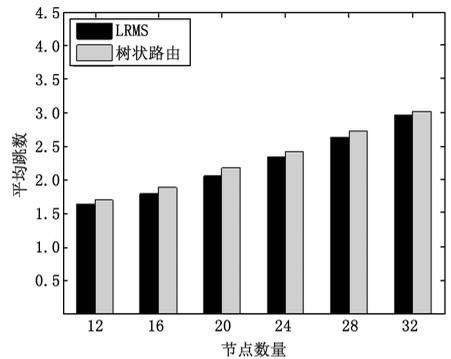


图 5 不同网络规模下的平均跳数

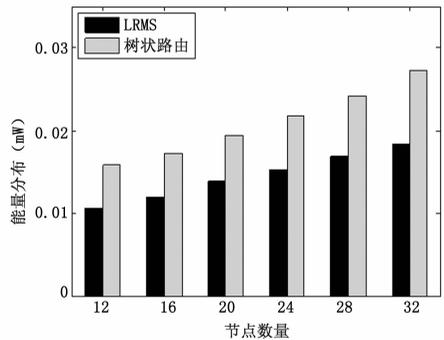


图 6 不同网络规模下的能量分布

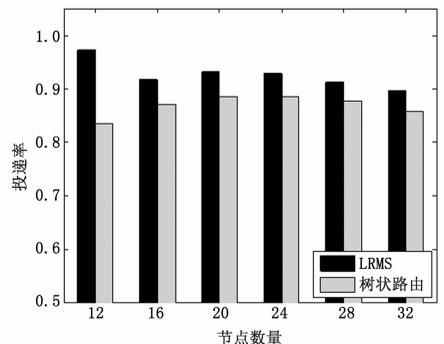


图 7 不同网络规模下的投递率

布以及投递率的对比。由图 5 可知，随着网络规模的增大，两种算法的平均跳数会成比例的增大，LRMS 具备更小的时延。图 6 说明，随着网络规模的增大，两种方法的能量分布均有所增大，LRMS 增长速度比树状路由小，且 LRMS 的能量分布

图 5~7 为不同网络规模下两种方法的平均跳数、能量分

总是低于树状路由; 说明 LRMS 能保证更为均衡的能耗和更长的网络生存时间。图 7 说明, 随着网络规模的增大, 两种方法的投递率均有所降低, LRMS 在不同的网络规模下均具备更高的投递率, 说明了数据传输的可靠性。综上, 基于 LRMS 的路由能够更好的适应变化的网络规模。

如图 8~10 所示为汇聚节点在不同速度下两种方法的平均跳数、能量分布以及投递率的对比。图 8 说明, 汇聚节点的速度变化对两种算法的平均跳数影响不大, LRMS 的平均跳数比树状路由的小。图 9 说明, 当汇聚节点速度小于 5 m/s 时, 两种算法的能量分布相差不大, 随着汇聚节点速度的增大, 树状路由的能量分布会急剧变坏, 而 LRMS 则受汇聚节点速度的影响较小, 说明 LRMS 具有更加均衡的能耗和更长的网络生存时间, 且能够更好的适应汇聚节点速度的增大。图 10 说明, 当汇聚节点速度小于 10 m/s 时, 树状路由的投递率比 LRMS 要好。随着汇聚节点速度进一步增大, 树状路由的投递率会急剧减小, 而 LRMS 则只有轻微的降低; 当汇聚节点速度超过

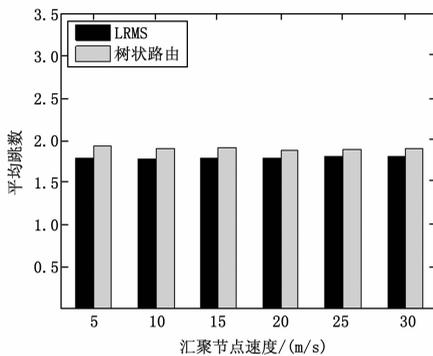


图 8 不同汇聚节点速度下的平均跳数

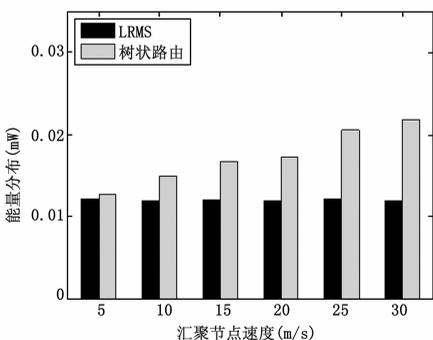


图 9 不同汇聚节点速度下的能量分布

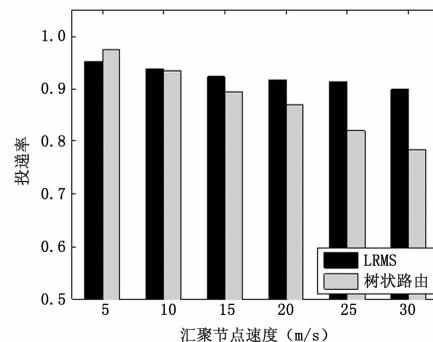


图 10 不同汇聚节点速度下的投递率

20 m/s 时, LRMS 的投递率相比树状路由具有非常明显的优势。因此, LRMS 可以更好地适应汇聚节点速度的增大, 更能适应由于汇聚节点快速移动带来的网络状况的快速变化。

4 结论

为实现无线传感器网络面向移动汇聚节点的数据传输, 本文提出一种基于强化学习的支持移动汇聚节点的自适应路由方法 (LRMS)。基于 Q 学习建立路由模型作为 LRMS 的基础, 设计了一种综合的奖励函数以对路由过程中的时延、投递率以及能耗均衡等多个性能指标进行优化。通过不断的学习网络状态, LRMS 可以本质上支持汇聚节点的移动性。理论分析表明 LRMS 具备快速收敛、低开销、存储计算需求低等特点, 适用于能量和资源受限的传感器节点。仿真结果表明, LRMS 能够保证包括时延、能耗均衡以及投递率在内的多个性能指标; 通过与树状路由的对比分析, 证明了 LRMS 方法的优越性。

参考文献:

- [1] Garca Villalba L J, Sandoval Orozco A L, Triviño Cabrera A, et al. Routing Protocols in Wireless Sensor Networks [J]. Sensors, 2009, 9 (11): 8399-8421.
- [2] Francesco M D, Das S K, Anastasi G. Data collection in wireless sensor networks with mobile elements: a survey [J]. ACM Transactions on Sensor Networks, 2011, 8 (1), 7: 1-31.
- [3] Tan C G, Xu K, Wang J X, et al. A sink moving scheme based on local residual energy of nodes in wireless sensor networks [J]. Journal of Central South University of Technology, 2009, 16 (2): 265-268.
- [4] Khan A W, Abdullah A H, Anisi M H, et al. A comprehensive study of data collection schemes using mobile sinks in wireless sensor networks [J]. Sensors, 2014, 14 (2): 2510-2548.
- [5] Carballido Villaverde B, Rea S, Pesch D. InRout - A QoS aware route selection algorithm for industrial wireless sensor networks [J]. Ad Hoc Networks, 2012, 10 (3): 458-478.
- [6] Luo H, Ye F, Cheng J, et al. TTDD: two-tier data dissemination in large-scale wireless sensor networks [J]. Wireless Networks, 2005, 11 (1-2): 161-175.
- [7] Konstantopoulos C, Pantziou G, Gavalas D, et al. A Rendezvous-Based Approach Enabling Energy-Efficient Sensory Data Collection with Mobile Sinks [J]. IEEE Transactions on Parallel and Distributed Systems, 2012, 23 (5): 809-817.
- [8] Li X, Yang J, Nayak A, et al. Localized Geographic Routing to a Mobile Sink with Guaranteed Delivery in Sensor Networks [J]. IEEE Journal on Selected Areas in Communications, 2012, 30 (9): 1719-1729.
- [9] Hasan A. A. Al-Rawi, Ming Ann Ng, Kok-Lim Alvin Yau. Application of reinforcement learning to routing in distributed wireless networks: a review [J]. Artificial Intelligence Review, 2013 (1): 1-36.
- [10] Boyan J A, Littman M L. Packet Routing in Dynamically Changing Networks: A Reinforcement Learning Approach [J]. Advances in Neural Information Processing Systems, 1994, 6: 671-678.
- [11] Kulkarni R V, Forster A, Venayagamoorthy G K. Computational Intelligence in Wireless Sensor Networks: A Survey [J]. IEEE communications surveys and tutorials, 2011, 13 (1): 68-96.
- [12] Guo W, Zhang W. A survey on intelligent routing protocols in wireless sensor networks [J]. Journal of Network and Computer Applications, Feb. 2014, 38: 185-201.