

A Best Clustering Scheme Based on Simulated Annealing Algorithm in Wireless Sensor Networks

LI Fangfang¹, WANG Jing^{2*}

(1. Electrical Engineering, Lanzhou Polytechnical College, Lanzhou 730050, China;
2. Institute of Communication Engineering, P. L. A University of Science and Technology, Nanjing 210007, China)

Abstract: Through the thorough study of existing clustering algorithms, with the problem of typical combinatorial optimization and the idea of simulated annealing algorithm, a best clustering scheme of the wireless sensor network was proposed based on simulated annealing algorithm. The assembled cluster head nodes were assembled and abstracted to a referenced model, objective cluster head collection was obtained through certain perturbation factors from continuous iteration. The idea of simulated annealing algorithm was used to the choosing process of the best number of clusters. This paper introduces the realization of the whole process in detail. The best number of clusters, persistent time, iterative times are studied and analyzed through established mathematics model and simulated experiments.

Key words: wireless sensor network; clustering; simulated annealing algorithm; the best number of clusters

EEACC: 6150P; 11101

doi: 10. 3969/j. issn. 1004-1699. 2011. 06. 024

一种基于模拟退火算法的无线传感器网络最优簇类求解方案

李芳芳¹, 王靖^{2*}

(1. 兰州工业高等专科学校电气工程系, 兰州, 730050;
2. 解放军理工大学通信工程学院, 南京, 210007)

摘要: 通过对现有分簇算法的深入研究, 从典型的组合优化问题着手, 利用模拟退火算法的思想, 提出了一种基于模拟退火算法的无线传感器网络最优簇类求解方案, 首先将簇头节点集合抽象为参考模型, 通过一定的扰动因子不断迭代更新最终得到目标簇头集, 然后将模拟退火算法的整个思想充分结合到最佳簇数的选择过程中, 详细介绍了整个过程的实现, 并通过数学建模分析和仿真实验相结合的方法对所给出方案中的最佳簇数、每轮持续时间、算法迭代次数等相关参数进行了研究和分析。

关键词: 无线传感器网络; 分簇; 模拟退火算法; 最佳簇数

中图分类号: TP301. 6

文献标识码: A

文章编号: 1004-1699(2011)06-0900-05

无线传感器网络^[1-3] (WSN, Wireless Sensor Network) 中, 节点数量多、体积小、能量有限, 处理能力、存储容量、通信能力都比较低, 如何充分利用这些能量和资源, 尽可能多地向数据收集者 (通常为基站 BS) 输出监视区域的有效数据, 成为一个热点研究问题。

将大型网络划分成若干个簇, 每个簇由一个簇首和若干成员节点组成, 实现层次化管理的分簇技术早已经应用在动态和自组织网络中, 但传感器网络在节点属性和优化目标等方面与上述网络有所不同, 需要重新研究相关算法。

无线传感器网络中, 成簇算法首先考虑的是宏

观上使功耗在所有传感器节点中进行公平合理的分配。实现最优的簇类划分是一种典型的组合优化问题, 可描述如下^[4]:

m 维空间上的 n 个模式 $\{X_i | i = 1, 2, \dots, n\}$, 要求聚成 k 类, 使得各类自身内的点间距离最近, 即要求:

$$x^2 = \sum_{i=1}^n \|X_i^{(p)} - R_p\| \quad (1)$$

达到最小, 其中 R_p 为第 p 类的中心:

$$R_p = \sum_{i=1}^n \frac{X_i^{(p)}}{n_p} \quad (2)$$

$p = 1, 2, \dots, k, n_p$ 为第 p 类中的点数。

簇类问题描述很简单,但最优化求解却很困难,主要原因是所谓的“组合爆炸”,簇类的可能划分方式有 $k^n/k!$ 个,要得到最优的簇类划分必定要运行复杂的算法,这在能量和计算能力极其有限的传感器节点上显然很难实现。

在无线传感器网络中,由于 Sink 节点的特殊性,它的存在意味着网络的存在,没有 Sink 节点,网络就没有存在的意义。相对于传感器节点而言, Sink 节点能源不受限制,运算能力和存储能力都很强,可以运行复杂的算法,而且对全网状况有着更全面的了解,这使得利用其生成最优的簇类成为可能。基于此,本文给出了一种由 Sink 节点运行模拟退火算法来得到最优簇类划分的求解方案。

1 模拟退火算法

模拟退火算法^[5] SA (Simulated Annealing) 源于统计物理学。根据统计热力学,物体中的每个分子的状态服从 Gibbs 分布,即:

$$\rho(r_i) = \frac{\exp\left(-\frac{E(r_i)}{kT}\right)}{\sum_{j=1}^M \exp\left(-\frac{E(r_j)}{kT}\right)} \quad (3)$$

式中: $E(r_i)$ 为第 i 个分子的能量函数; r_i 为第 i 个分子所处的状态; k 为波耳兹曼常数, T 表示温度; $\rho(r_i)$ 为第 i 个分子的概率密度。

模拟退火算法最早由 Metropolis 在 1953 年提出,由 Kirkpatrick 于 1983 年成功地应用在组合优化问题中,目前已经应用到各门学科中以解决非线性系统中的优化问题。算法源于对实际固体退火过程的模拟,即先将固体加温至充分高,再逐渐冷却。加温时,固体内部粒子变为无序状态,内能增大;而逐渐降温时,粒子趋于有序,在每个温度都达到平衡态,最后在常温时达到基态,内能减为最小。因此,算法实际上是将优化问题类比为退火过程中能量的最低状态,也就是温度达到最低点时,概率分布中具有最大概率的状态。

2 模拟退火算法的最优簇类求解方案

当前,无线传感器网络最优簇类的求解问题是一个著名的 NP 难题^[6],本文在参考其他研究人员在这方面研究成果^[7-12]的基础上,设计了一种基于“轮”的思想的网络,其采用模拟退火算法来得到簇类划分的近似最优解(为了简单起见,本文称之为最优解)。

2.1 算法的体系结构

网络在整个生命周期内划分为若干“轮”,在新

一轮开始时, Sink 节点首先对当前存活的节点数进行统计,并据此计算本轮的最优簇数和全网平均能量,高于全网平均能量的节点加入到簇头节点的集合,反之,则加入到非簇头节点的集合。接着, Sink 节点对簇头节点的集合运行模拟退火算法进行分簇,在每个簇内选择一个剩余能量最多的普通节点作为该簇的备用簇头节点,然后向全网广播簇划分情况,各节点按 Sink 节点规定的时隙与所属簇的簇头节点进行通信。本轮工作结束后,进入下一轮操作,并重复此过程,如图 1 所示。

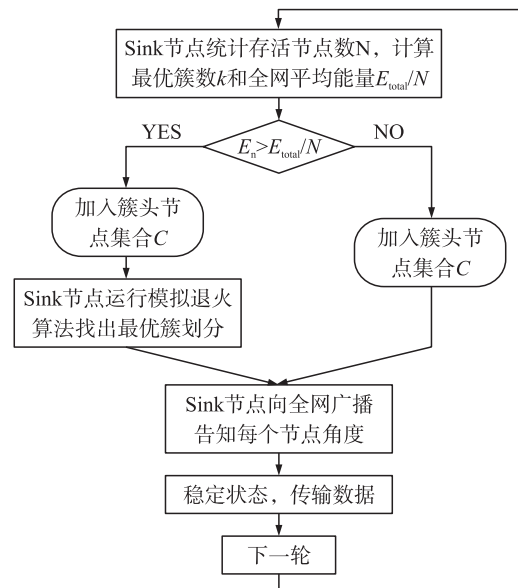


图1 网络运行流程

簇头节点保证在当前轮中整个网络的能耗最低。整个网络的能耗最低是指所有簇类的普通节点向其簇头节点传输数据所耗费的能量总值最小,这个值我们通过计算簇内普通节点和簇头节点之间的最小距离来决定。实际上,节点在通信过程中所耗费的能源与节点间距离并不是非常精确的正比关系,但是,收集所有节点间的信道信息并不现实,因此,使用距离作为能源损耗的近似值。本文使用第一顺序无线电能量模型^[7] (First Order Radio Model) 做为能量模型。

成簇算法的具体流程如图 2 所示:在每次迭代过程中,当前状态的簇头节点集合 C ,通过随机且独立的扰动 C 中的节点 c 的坐标 (x, y) 得到其新坐标 (x', y') ,位置最靠近 (x', y') 的节点就成为新集合 C' 中的 c' 。在给定的迭代值 k 下,包含节点 c 的集合 C 的代价为 $f(C)$,包含节点 c' 的集合 C' 的代价为 $f(C')$,如果 $f(C') < f(C)$,则接收 C' 为下一状态的簇头集;否则,将依照概率 P_k 判断 C' 是否为下一状态的簇头集, P_k 表示如下:

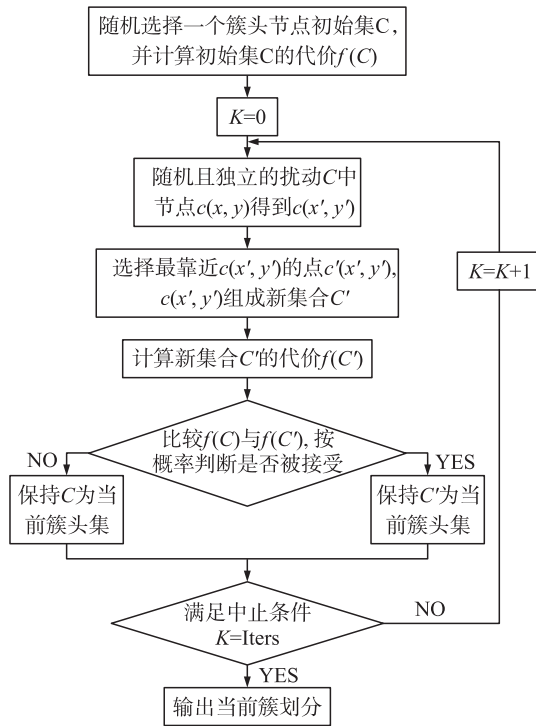


图2 簇类形成算法

$$P_k = \begin{cases} e^{-\frac{f(C')-f(C)}{\alpha_k}} & f(C') \geq f(C) \\ 1 & f(C') < f(C) \end{cases} \quad (4)$$

$f(C)$ 代表代价, 定义如下:

$$f(C) = \sum_{i=1}^N \min_{c \in C} d^2(i, c) \quad (5)$$

$d(i, c)$ 表示节点 i 和节点 c 之间的距离。

α_k 是控制参数, α_k 必须随同 k 一起增长以保证算法的收敛, α_k 定义如下:

$$\alpha_k = 1000e^{-\frac{k}{20}} \quad (6)$$

算法究竟要经过多少次迭代才会收敛, 取决于具体的网络环境, 通常用试验的方法找出迭代次数的值来确定算法何时终止, 后文将针对具体的仿真环境对迭代次数做进一步的研究。

2.2 算法相关参数的建模分析

2.2.1 最佳簇类个数

节点在工作过程中会因为故障或能量耗尽而死亡, 能够参与工作的节点总数随着网络工作时间的增长在不断的减少, 本文提出一种基于当前存活节点数的最佳簇数的确定机制:

假设当前网络中存活的节点数为 N_r , 其随机分布在一块 $M \times 2M$ 的矩形区域, 如果有 k 个簇, 那么每簇有 N_r/k 个节点 (每簇有一个簇头和 $(N_r/k) - 1$ 个簇成员)。簇头节点的能量消耗可分为三部分: 接收从簇内普通节点传输的数据, 进行数据融合, 将融合后的数据向 Sink 节点传输。因此, 簇头传输一

帧数据所消耗的能量为:

$$E_{CH} = lE_{elec} \frac{N_r}{k} + lE_{DA} \frac{N_r}{k} + l\epsilon_{mp} d_{toSink}^4 \quad (7)$$

l 表示每个节点一帧内的数据包的特位数。

每一个簇成员节点在每一帧中只需向簇头节点传输数据一次, 那么, 每一个簇成员节点消耗的能量为:

$$E_{non-CH} = lE_{elec} + l\epsilon_{friss} d_{toCH}^2 \quad (8)$$

每个簇所占据的面积大约为 $2M^2/k$, 这是一个节点以 $\rho(x, y)$ 的密度分布的任意形状的区域。假设簇头位于簇的中间, 那么, 从簇成员节点到簇头节点的平方距离可以通过下式计算:

$$d_{toCH}^2 = \iint (x^2 + y^2) \rho(r, \theta) r dr d\theta \quad (9)$$

为了便于研究, 假设这个区域是圆形的, 则圆半径 $R = \sqrt{2M}/\sqrt{k\pi}$, $\rho(r, \theta)$ 对 r 和 θ 是恒定的, 上式就简化为:

$$d_{toCH}^2 = \rho \int_{\theta=0}^{2\pi} \int_{r=0}^{\sqrt{2M}/\sqrt{k\pi}} r^3 dr d\theta = \frac{2\rho}{\pi} \frac{M^4}{k^2} \quad (10)$$

如果簇内节点的密度是均匀的, 那么 $\rho = \frac{1}{2M^2/k}$, 则:

$$d_{toCH}^2 = \frac{1}{\pi} \frac{M^2}{k} \quad (11)$$

将式(11)代入式(8)可得:

$$E_{non-CH} = lE_{elec} + l\epsilon_{friss} \frac{1}{\pi} \frac{M^2}{k} \quad (12)$$

由式(7)和式(12)可得在一帧中一个簇的能量消耗为:

$$E_{cluster} = E_{CH} + \left(\frac{N_r}{k} - 1\right) E_{non-CH} \approx E_{CH} + \frac{N_r}{k} E_{non-CH} \quad (13)$$

在一帧中所有 N_r 个节点消耗的能量为:

$$E_{total} = kE_{cluster} = l \left(E_{elec} N_r + E_{DA} N_r + k\epsilon_{mp} d_{toSink}^4 + E_{elec} N_r + \epsilon_{friss} \frac{1}{\pi} \frac{M^2}{k} N_r \right)$$

此时, 将 E_{total} 对 k 求导并让其等于零, 可求得理想的簇头数 k , 即:

$$\frac{dE_{total}}{dk} = 0 \quad (14)$$

$$\epsilon_{mp} d_{toSink}^4 = N_r \epsilon_{friss} \left(\frac{1}{\pi} \frac{M^2}{k^2} \right) \quad (15)$$

在上式中, d_{toSink} 在具体网络中可给出其确定的上下边界值, k 和 N_r 则是两个不确定的变量, 如果以 k_{opt} 表示最佳的簇头节点个数, 对式(15)整理后得:

$$\frac{k_{opt}}{\sqrt{N_r}} = \frac{1}{\sqrt{\pi}} \sqrt{\frac{\varepsilon_{friss}}{\varepsilon_{mp}} \frac{M}{d_{toSink}^2}} \quad (16)$$

在新一轮开始时, Sink 节点首先对当前仍存活的节点数 N_r 进行统计, 并采用式(16)计算本轮的最优簇数。这里 $k_{opt}/\sqrt{N_r}$ 只是给出了一个最优的取值范围, 后文将针对具体的网络环境给出进一步的仿真分析。

2.2.2 每轮持续时间

选择合适的每轮持续时间至关重要, 如果一轮持续时间太短, 则网络处于不断的动态划分过程, 会造成能源的浪费, 如果一个回合持续时间太长, 又会使簇头节点很快耗尽能量, 甚至在下一回合中无法充当普通节点的角色。

假设网络的初始节点数为 N , 并且每个节点的能源能够持续足够长的时间允许在它的生命周期中成为一次簇头节点和在其他回合中成为普通节点。如果整个网络生命周期共有 N/k 轮, 这就意味着每个节点有足够的能源在某一轮担当一次簇头节点, 在其他的 $(N/k) - 1$ 轮中则为普通节点。如果以 E_{start} 表示每个节点的初始能量, 则:

$$E_{CH} + \left(\frac{N}{k} - 1\right) E_{non_CH} = E_{start} \quad (17)$$

假设每轮 TDMA 帧的个数为 S_T , 将式(7)和式(8)代入式(17)可得:

$$S_T \left(l E_{elec} \frac{N}{k} + l E_{DA} \frac{N}{k} + l \varepsilon_{mp} d_{toSink}^4 \right) + S_T \left(\frac{N}{k} - 1 \right) \left(l E_{elec} + l \varepsilon_{friss} \frac{1}{\pi} \frac{M^2}{k} \right) = E_{start}$$

因此, S_T 可表示为:

$$S_T = \frac{E_{start}}{l \left[\left(E_{elec} \frac{N}{k} + E_{DA} \frac{N}{k} + \varepsilon_{mp} d_{toSink}^4 \right) + \left(\frac{N}{k} - 1 \right) \left(E_{elec} + \varepsilon_{friss} \frac{1}{\pi} \frac{M^2}{k} \right) \right]}$$

假设每簇有 N/k 个节点, 数据传输速率为 R_b , 则 l bit 数据持续时间为 $t_m = \frac{l}{R_b}$, 一个 TDMA 帧持续

时间为 $t_T = \frac{N}{k} \frac{l}{R_b}$, 那么每轮的持续时长为:

$$t_r = S_T \times t_T = \frac{N}{k R_b} \times$$

$$\frac{E_{start}}{l \left[\left(E_{elec} \frac{N}{k} + E_{DA} \frac{N}{k} + \varepsilon_{mp} d_{toSink}^4 \right) + \left(\frac{N}{k} - 1 \right) \left(E_{elec} + \varepsilon_{friss} \frac{1}{\pi} \frac{M^2}{k} \right) \right]}$$

2.3 特定网络模型中相关参数的仿真分析

考虑一组传感器节点随机散布于监测区域的情形: 网络大小为 $100 \text{ m} \times 200 \text{ m}$, 随机分布着 100 个传

感器节点, 传感器节点的初始能量为 2 J, 采取连续发送模式, 每个节点每个时隙传输 4 000 bit (500 byte) 的感知数据, Sink 节点坐标为 $\langle 50, 185 \rangle$ 。

2.3.1 簇头数目的选择

上文式(16)得出的只是对理想的簇头数目一般性的粗略估计值, 针对所给网络模型, 对其最佳取值做进一步的分析。从上文可知, $M = 100 \text{ m}$, $100 < d_{toSink} < 238$, 网络初始节点数为 100, 将上述数值代入式(16)可得:

$$0.1 < k_{opt}/\sqrt{N_r} < 0.6 \quad (18)$$

在第一轮, 所有节点都能量充足, 假设都可以参与工作, 即 $N_r = 100$, 将其代入上式可得:

$$1 < k_{opt} < 6 \quad (19)$$

对以上两者的取值分别进行仿真, 仿真结果如图3和图4所示。图3考察了第一轮簇头节点数 k 取不同值时, 网络在本轮消耗的总能量的变化情况, 从仿真结果可以看出, 当 k 取 5 时本轮消耗的能量达到最小; 图4考察了当 $k_{opt}/\sqrt{N_r}$ 取不同比值时, 整个网络生命周期内平均每轮消耗能量的变化情况, 可以看出, 比值为 0.4 时平均能耗最小。进一步的分析可得, 当簇头节点太少时, 簇成员要传送数据给距离很远的簇头节点, 数据融合也较少, 因此需要传输较大的数据量, 从而消耗了较多的能量; 当簇头数目过多时, 数据融合的开销增加, 而实际上没必要那么多簇头进行数据融合的工作, 这样也将导致消耗额外的能量; 而且, 随着网络工作时间的增加, 由于存活节点会不断减少, 簇头节点应该呈减少趋势。



图3 第一轮中簇头节点数与能量消耗

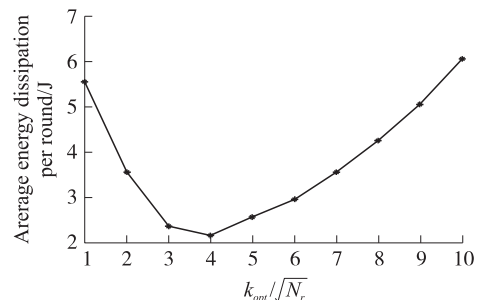


图4 $k_{opt}/\sqrt{N_r}$ 比值与能量消耗

2.3.2 模拟退火算法迭代次数的确定

文中由 Sink 节点运行模拟退火算法来决定下一轮 k_{opt} 个最优簇的簇头节点由哪些节点担任,而算法究竟要经过多少次迭代才会收敛,要取决于具体的网络环境。

为确定出本文所给网络模型中算法要经过多少次迭代才可终止,我们在网络中有 100 个节点的情况下(存活节点最多时),对算法进行了仿真,仿真结果如图 5 所示。图中横轴表示迭代次数,纵轴表示当前簇划分状态下网络通信的最小代价,算法在 200 次迭代以后开始收敛。由于这些计算是在 Sink 节点进行的,所以其运算代价并不是我们所关心的。

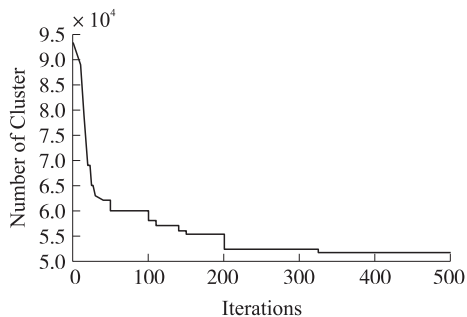
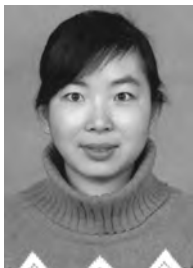


图5 迭代次数与网络通信代价

3 结束语

本文以无线传感器网络为研究背景,对最优簇类确定机制进行了深入研究,给出了一种基于模拟退火算法的无线传感器网络最优簇类确定方案,引入最优簇数可根据网络状况动态变化的簇类个数选择机制,以使网络在较低的能耗水平下获得更为合理的簇类划分,并通过数学建模分析和仿真实验相结合的方法对该方案中的相关参数进行了研究和确定,研究在簇头数据处理阶段如何具体地进行数据处理,如何高效可靠的传输数据将是下一步的主要工作。



李芳芳(1981-),女,甘肃省兰州市人,讲师,硕士研究生,2004年毕业于甘肃工业大学,2008年至今在兰州理工大学攻读硕士学位,主要从事传感器技术方面的研究;

参考文献:

- [1] Tilak S, Abu-Ghazaleh N B, Heinzelman W. A Taxonomy of Wireless Micro-Sensor Network Models[J]. Mobile Computing and Communications Review, 2002, 1(2): 1-8.
- [2] Arampatzis T H, Lygeros J, Manesis S. A Survey of Applications of Wireless Sensor Networks[C]//Proceedings of the 13th Mediterranean Conference on Control and Automation. Limassol; Cyprus, 2005: 27-29.
- [3] 陈闻杰, 陈迅, 高丽强, 等. 无线传感器网络成簇算法研究[J]. 小型微型计算机系统, 2008, 29(2): 219-225.
- [4] 王春. 无线传感器网络路由协议的设计与仿真[D]. [硕士学位论文]. 成都: 电子科技大学. 2004. 47-48.
- [5] 蒋龙聪, 刘江平. 模拟退火算法及其改进[J]. 工程地球物理学报, 2007, 4(2): 135-140.
- [6] Agarwal P, Procopiuc C. Exact and Approximation Algorithms for Clustering[C]//Proceedings of the Ninth Annual ACM-SIAM Symposium on Discrete Algorithms, 1999, 658-667.
- [7] Wendi Beth Heinzelman. Application-Specific Protocol Architectures for Wireless Networks[D]. [Ph. D. Thesis]. 77 Massachusetts Avenue, Cambridge, MA 02139, U. S. A. : Electrical Engineering and Computer Science at the Massachusetts Institute of Technology, June 2000.
- [8] Heinzelman W, Chandrakasan A, Balakrishnan H. An Application Specific Protocol Architecture for Wireless Microsensor Networks[J]. IEEE Transactions on Wireless Communications, 2002, 4(1): 660 - 670.
- [9] Jun Lu, Tatsuya Suda. Coverage-Aware Self-Scheduling in Sensor Networks[C]//Computer Communications, 2003, CCW 2003, Proceedings. 2003 IEEE 18th Annual Workshop on 20-21 Oct. 2003, 117-123.
- [10] Juhana Yrjölä. Energy-Efficient Communication Protocol for Wireless Microsensor Networks[R]. T-79. 194 Seminar on Theoretical Computer Science 2005 Algorithmics of Sensor Networks.
- [11] 吴砥柱, 金心宇, 张昱, 等. 基于投票策略的异构传感器网络分簇算法[J]. 传感技术学报, 2009, 22(4): 537-542.
- [12] 李田, 史浩山, 杨俊刚. 无线传感器网络 LEACH 协议成簇算法研究[J]. 传感技术学报, 2010, 23(8): 1158-1162.



王靖(1980-),男,甘肃省兰州市人,博士研究生,2008年至今在解放军理工大学攻读博士学位,主要从事传感器网络方面的研究。