

# 面向分簇的 WSN 数据融合在瓦斯监控中的应用<sup>①</sup>

张长森, 闵 杰

(河南理工大学 计算机科学与技术学院, 焦作 454000)

**摘 要:** 为了更好地对矿井瓦斯进行监控, 在迅速提取异常瓦斯数据的同时消除非异常数据带来的冗余, 本文提出了一种基于分簇的 WSN(无线传感器网络)数据融合方法. 它利用总能耗最小的方法将所有节点最佳分簇, 每个簇的节点通过竞选簇头概率的大小来确定簇头, 非簇头节点通过两次阈值判断采集的数据是否异常并将异常数据发送到监控基站供工作人员决策. 实验仿真表明: 该方法延长了整个传感器网络的生存周期, 对矿井瓦斯起到有效监控.

**关键词:** 矿井; 异常瓦斯数据; WSN; 数据融合; 最佳分簇; 生存周期; 监控

## WSN Data Fusion Based on Cluster in Mine Gas Monitoring Application

ZHANG Chang-Sen, MIN Jie

(Computer Science and Technology Institute, Henan Polytechnic University, Jiaozuo 454000, China)

**Abstract:** To monitor mine gas better, extract the abnormal gas data rapidly and eliminate the normal data redundancy, this paper proposes a WSN data fusion based on cluster method. It used the way of minimizing the total energy consumption that all nodes are divided into clusters best, nodes in each cluster determined the cluster head through the election of cluster head probability, the cluster nodes judged the data collected whether to be abnormal or not through the two thresholds and sent it to the monitoring stations for staffs to decide. The simulation results show that this method prolongs the wireless sensor network lifetime, the mine gas is monitored effectively.

**Key words:** mine; abnormal gas data; wsn; data fusion; best clustering; lifetime; monitoring

### 1 引言

随着矿井瓦斯事故的频繁发生, 有必要采取一种有效的通信方式对矿井瓦斯进行监控. 目前绝大多数矿井采用有线的通信方式, 但是有线通信存在布线困难、维护难度大等一些限制性因素, 这就需要采取更为方便有效的无线传感器网络 (WSN, Wireless Sensor Network). WSN 近年来成为无线通信领域的一个热门研究话题. 它是通过一系列无线传感器节点, 以自组织的形式连接到一起而形成的网络. 但是由于节点采用一般的电池供电, 能量有限, 不适于传输大量的数据, 所以要对大量数据进行融合, 减少网内数据的传输. 目前常用的数据融合方法有: 加权融合法, D-S 理论法, 模糊算法, 贝叶斯估计法, 卡尔曼滤波法等. 文献[1]提出了一种基于树的数据融合算法, 该算法是将

所有节点分成父节点和子节点, 通过将所有子节点采集的数据中相同的数据除去, 然后传输到 sink 节点, 这种方法可以去除一定的冗余. 但是在实际矿井环境中, 节点采集的数据大部分是不相同的, 而是相近的, 这就不能最大限度地去除冗余数据, 降低节点能耗. 文献[2]提出了一种基于克隆选择的数据聚合算法, 采用最小生成树分析网络抗体间的关系, 计算抗体间的亲和力, 亲和力低于一定阈值要进行一定比例的清除, 这样可以消除冗余数据, 但是相关的参数设置和阈值很难把握. 文献[3]通过对所有簇内节点采集的数据自适应加权平均计算, 然后发送给 sink 节点, 虽然这种方法能够一定程度地降低数据冗余, 但是在矿井瓦斯监测中不能提取出异常瓦斯数据. 针对这些情况, 本文提出了面向分簇的 WSN 数据融合方法. 它的思想

<sup>①</sup> 基金项目: 国家自然科学基金(51174263)

收稿时间: 2013-03-06; 收到修改稿时间: 2013-03-25

是首先对矿井工作面区域内所有传感器网络节点分簇, 然后通过使所有节点的总能耗最小计算出最佳的分簇个数, 这样就对所监控的区域进行合理划分. 然后按照簇内节点竞选簇头的概率大小选出每个簇的簇头节点, 以均衡簇内节点的能耗. 之后对于节点采集的原始瓦斯数据值通过两次阈值算法确定原始数据是否发送到监控基站. 第一次设定的阈值是过滤那些相似的冗余数据, 第二次设定的阈值是确定超过第一次阈值范围的数据值是否是异常数据, 这样既可以降低节点传输数据带来的能量消耗, 又可以检测出矿井环境中的异常瓦斯数据, 使地面工作人员能够有效地对矿井环境实施监控.

### 1 无线传感器网络模型及节点能耗分析

将  $n$  个传感器节点以密度  $s$  随机均匀的分布在一个半径为  $R$  的圆形区域内, 且传感器网络具有以下性质:

- 1) 传感器网络中每个节点都是固定的,  $sink$  节点位于传感器网络覆盖区域之外.
- 2) 每个节点的初始能量相同, 数据采集周期为  $T$ , 每次采集的数据量为  $l$ bits.
- 3) 每个节点均具有数据处理能力, 能将采集的原始瓦斯数据或者所有来自簇内成员节点的长度为  $l$ bits 的数据包压缩为一个长度为  $l$ bits 的数据包, 然后发送给监控基站.

#### 1.1 网络能耗模型<sup>[4]</sup>

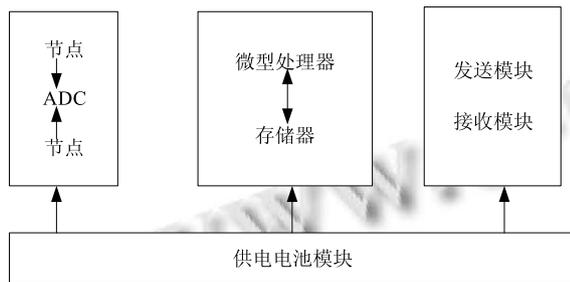


图 1 无线传感器网络能耗模型

设每个节点发送数据包所需的能量为  $E_{TX}(l, d)$ , 接收数据包所需能量为  $E_{RX}(l)$ , 发射电路消耗能量为  $E_{elec}$ ,  $\sigma_1, \sigma_2$  为放大器系数,  $d_0$  为发送节点与接受节点之间的距离的临界值, 则:

$$E_{TX}(l, d) = E_{TX-elec}(l) + E_{TX-amp}(l, d) = \begin{cases} lE_{elec} + l\sigma_1 d^2 & (d \leq d_0) \\ lE_{elec} + l\sigma_2 d^2 & (d > d_0) \end{cases} \quad (1)$$

$$E_{RX}(l) = E_{RX-elec}(l) = lE_{elec} \quad (2)$$

其中,  $E_{elec}$  表示无线收发电路发送 1bit 数据所消耗的能量,  $\sigma_1 d^2$ ,  $\sigma_2 d^2$  分别表示自由空间模型和多径衰减模型中发送电路发送 1bit 数据所消耗的能量.

### 2 基于分簇的数据融合算法

#### 2.1 最佳分簇算法

将矿井工作面圆形区域内的  $n$  个瓦斯传感器节点均匀分成  $m$  个簇 (如图 2 所示), 簇内节点与簇头之间距离较近, 服从自由空间传播模型; 簇头和基站之间相距较远, 服从多径衰减传播模型. 簇内节点到簇头的距离记为  $d_{toCH}$ , 簇头到监控  $sink$  节点的距离记为  $d_{to sink}$ .

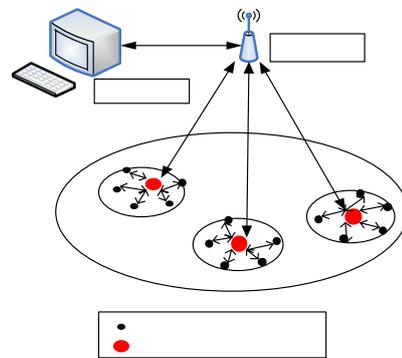


图 2 瓦斯传感器网络的拓扑结构

则每个簇成员节点能量消耗为:

$$E_{node} = lE_{elec} + l\sigma_1 d_{toCH}^2 \quad (3)$$

由于所有传感器节点以密度  $s$  均匀分布在半径为  $R$  的圆形区域内, 簇的大小基本相同, 所以每个簇面积的大小为  $\pi R^2/m$ , 每个簇可以近似看作一个圆, 其半径为  $R/\sqrt{m}$ . 所以

$$s = \frac{m}{\pi R^2} \quad (4)$$

$$E(d_{toCH}^2) = \iint (x^2 + y^2) s dx dy = s \int_0^{2\pi} \int_0^{R/\sqrt{m}} r^2 * r dr d\theta = \frac{R^2}{2m} \quad (5)$$

这里用(5)式代替(3)式中的  $d_{toCH}^2$ , 得到:

$$E_{node} = lE_{elec} + \frac{R^2}{2m} l\sigma_1 \quad (6)$$

当簇头接收到所有来自簇内节点传输的长度为  $l$  的瓦斯数据时, 要将它们压缩成长度为  $l$  的数据包, 然后经过簇头节点和  $sink$  节点发送到监控基站, 设数据处理消耗的能量为  $E_{DF}$ , 则簇头对来自簇内  $(n/m-1)$  个节

点的数据进行处理的能耗为:

$$E(l, n) = n l E_{DF} \quad (7)$$

则簇头节点的能耗为:

$$E_{CH} = l E_{elec} \left( \frac{n}{m} - 1 \right) + l E_{DF} \frac{n}{m} + l E_{elec} + l \sigma_2 d_{to\ sink}^4 \quad (8)$$

所以整个网络传输一次瓦斯数据所消耗的能量为:

$$E_{WSN} = (n - m) E_{node} + m E_{CH} = (n - m) \left( l E_{elec} + l \sigma_1 \frac{R^2}{2m} \right) + m \left[ l E_{elec} \left( \frac{n}{m} - 1 \right) + l E_{DF} \frac{n}{m} + l E_{elec} + l \sigma_2 d_{to\ sink}^4 \right] \quad (9)$$

使  $E_{WSN}$  最小, 可得:

$$m_{opt} = \sqrt{\frac{n \sigma_1 R^2}{2 \sigma_2 d_{to\ sink}^4 - 2 E_{elec}}} \quad (10)$$

由文献[5]知:

$$d_{to\ sink} = \frac{0.765 \sqrt{\pi} R}{2} \quad (11)$$

把(11)式代入(10)式得:

$$m_{opt} = \sqrt{\frac{n \sigma_1 R^2}{0.042 \sigma_2 \pi^2 R^4 - 2 E_{elec}}} \quad (12)$$

求出每个簇的节点个数之后, 由于在实际数据传输过程中采集节点距离 *sink* 节点很远, 不可能让每一个节点采集的数据都分别传送到 *sink* 节点, 这样带来的能量消耗是很大的. 这样就需要在每一个簇内选择一个簇头节点负责簇内节点到 *sink* 节点数据的转发, 下面介绍簇头节点的选择算法.

### 2.2 簇头节点的选择算法

簇头节点不仅能够转发节点采集的瓦斯数据, 还可以均衡簇内节点能耗. 经典的 *LEACH* 协议<sup>[6]</sup>提出了通过随机生成数与设置阈值  $T(n)$  进行比较的方法随机选择簇头, 虽然一定程度上缓解了节点的能量消耗不均匀, 但是没有考虑节点的剩余能量对其成为簇头的影响. 文献[7]提出了一种基于剩余能量的簇头选择算法, 该方法通过节点的剩余能量确定节点成为簇头的概率大小, 但是没有考虑其他因素的影响, 比如簇头到节点的距离和簇头到 *sink* 节点的距离等等. 文献[8]提出了一种考虑了剩余能量、节点与簇头和 *sink* 之间距离这些参数来判定哪一个节点当选簇头的概率最大的簇头选择算法, 这种方法没有考虑到节点当选簇头的次数对后来当选概率大小的影响.

针对这种情况, 本文提出了一种既考虑到节点剩

余能量, 又兼顾了被选为簇头的节点到 *sink* 节点的距离及其到簇内节点的距离和当选簇头的次数的影响的簇头选择算法, 算法如下:

首先设节点的初始能量、剩余能量分别为  $E_{ini}$ ,  $E_{rem}$ , 令  $p_1 = E_{rem} / E_{ini}$ . 由于节点传输数据比处理数据消耗的能量大得多, 所以在选择簇头时我们考虑了传输距离  $d_{toCH}$ 、 $d_{toBS}$  对节点能耗的影响, 令  $p_2 = d_{toCH} / d_{toBS}$ . 当一个节点被选择为簇头节点后, 下一次当选的概率就会降低, 设节点被选为簇头的次数为  $N_i$  ( $i=1, 2, \dots, m$ ), 令  $p_3 = 1 / N_i$ . 为了避免被选为簇头的节点过早死亡, 在进行簇头选择时, 合理地分别给  $p_1, p_2, p_3$  一个权值  $k_1, k_2, k_3$ , 其中  $k_1 + k_2 + k_3 = 1$ . 设每个节点当选为簇头的概率为  $P$ , 则  $P = k_1 p_1 + k_2 p_2 + k_3 p_3$ .

对于如何确定每个簇中距离监控基站最近的节点, 我们需要监控基站向每个簇的所有成员节点发送一个“问候”数据包, 每个成员节点接收到该数据包之后要迅速返回一个“答复”, 基站根据各个节点接收和返回数据包的时间差可以知道对应节点到基站的距离. 假设各个簇与基站的通信互不干扰, 基站接收到的对应到每个簇的节点的答复数据包的时间最短, 说明该节点距离基站最近. 然后第一次我们选择每个簇中最接近监控基站的节点为簇头, 在每个簇内, 监控基站给各个簇内节点发送一个簇头重新选择周期, 该周期之后簇内节点根据自身成为簇头的概率大小重新竞争成为簇头节点. 节点会存储自身的剩余能量  $E_{ini}$  和传输距离  $d_{toCH}$ 、 $d_{toBS}$  信息, 节点当选簇头的次数初始值为 0, 每个节点会跟簇内其他节点逐次比较当选概率的大小, 每一次选取较大的概率  $p_i$ , 然后用  $p_i$  去和其他概率值比较, 直到比较完成得到最大的概率值, 其所对应的节点即为簇头节点. 簇头会向邻居节点广播当选消息, 每当选一次, 簇头节点本身要累计计数一次, 当选的次数  $N$  要暂存在缓存器中供下一次当选簇头与否做参考.

### 2.3 数据融合方法及异常瓦斯数据的识别

本文为了更加有效地识别矿井瓦斯异常数据, 提出了一种基于估计机制的数据融合方法, 在消除数据冗余的同时能够排除误采瓦斯数据的干扰.

首先, *WSN* 瓦斯传感器节点将采取的稳定的瓦斯浓度数据传输到监控基站, 基站工作人员根据所采集的正常数据设定一个瓦斯浓度变化阈值  $\theta_0$ , 然后传输给每一个节点, 节点保存到存储器中. 之后各簇内节

点在矿井工作面采集瓦斯浓度数据时,如果本次采集的数据值与前一次瓦斯浓度数据值变化量不大于于 $\theta_0$ ,即: $|\theta_i - \theta_{i-1}| \leq \theta_0$ ,节点处理器判别其为正常数据,不发送到簇头节点,减少了网内数据的传输量.当 $|\theta_i - \theta_{i-1}| > \theta_0$ 时,判断其可能为异常瓦斯数据,需要将本次采集的数据与前一次数据都暂时存放到节点缓存器中,等待一定周期 $T_1(T_1=2T$ ,其中 $T$ 为传感器节点的采集周期).在此期间,节点缓存器每收到一个可能的异常数据就要计数一次,同时节点要通知簇头节点,簇头节点会启动一个特定的传输通道使异常数据能够快速上传至 sink 节点.如果节点缓存器连续 $N(N>1)$ 次收到超过阈值范围的数据,就要将这 $N$ 个瓦斯数据传输到簇头节点,最后送至 sink 节点和监控基站,工作人员就会辨别井下瓦斯浓度是否超标(根据《煤矿安全规程》知:正常瓦斯浓度范围为 0-1%).否则,视前一次采集的数据为误采数据,这样就避免了误采数据的干扰,同时丢弃误采数据以减少节点的存储量.下一个周期采集的数据,节点要与异常数据之前的数据进行比较,依此达到循环监测.如果节点采集到的数据值大于 1%,则迅速通过簇头和 sink 节点传输到监控基站,工作人员及时采取措施减少瓦斯异常事故的发生.

### 3 仿真分析

本次实验仿真采用的数据如下:节点个数 $n=200$ ,圆形区域半径 $R=100m$ ,节点初始能量 $E_{ini}=5J$ ,发射电路消耗能量 $E_{elec}=10nJ/bit$ ,数据包长度 $l=500bits$ ,放大器系数 $\sigma_1, \sigma_2$ 分别为 $10pJ/bit/m^2, 0.0013pJ/bit/m^4$ ,数据处理消耗能量 $E_{DF}=5 nJ/bit/signal$ ,权值 $k_1, k_2, k_3$ 分别为 $0.5, 0.3, 0.2$ .

#### 3.1 对最佳分簇个数的仿真

将仿真设定的数据代入公式(11)计算得到最佳的分簇个数是 23.8(即 24)个,而图 3 仿真得到的最佳分簇个数为 20 到 25 之间,说明理论与实际情况大致相符合.

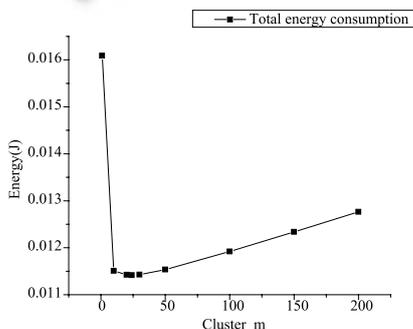


图 3 最佳分簇个数仿真

#### 3.2 对簇头选择算法存活节点个数的仿真

本文簇头选择算法将与 LEACH 进行网络的存活节点个数的比较,仿真图如下:

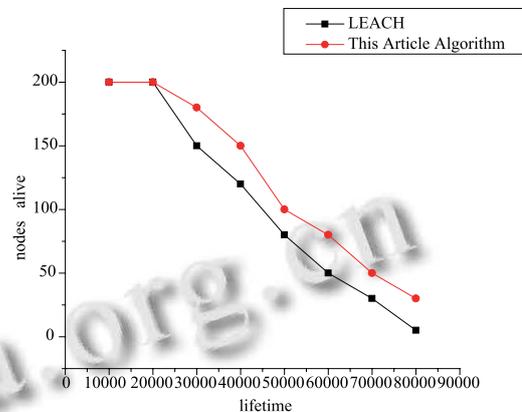


图 4 本文簇头选择算法与 LEACH 节点存活个数比较

由图 4 可知:在一定循环周期内本文簇头选择算法和 LEACH 算法中节点都处于存活状态,一定周期之后,LEACH 算法的第一个节点开始死亡且其死亡的时间早于本文算法,另外 LEACH 算法的节点死亡速率更快,说明本文算法能有效均衡节点能量,延长网络生存周期.

#### 3.3 对本文融合算法和加权均值算法以及未进行融合的总能耗比较的仿真

由于同一矿井工作面同一周期内采集的瓦斯数据存很多都相近(符合正态分布),这就需要消除相同的或者相近的数据.对本文算法中取初始阈值 $\theta_0=0.002$ ,本次采集的数据如果比前一次变化量超过 0.002,则要进行等待一个周期,判断是异常数据还是误采数据.节点每个周期采集的原始数据如图 5 所示:

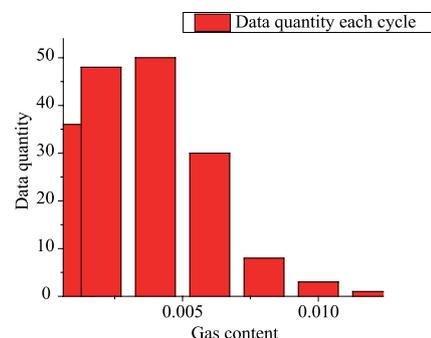


图 5 节点每个周期瓦斯浓度原始数据

对于图 5 的瓦斯数据图可知:每一轮(两个周期)每个簇的节点采集的瓦斯数据是异常数据的数量约为

4/bits,每个周期采集的数据量为 200/bits. 三种算法中整个网络的能耗如图 6 所示:

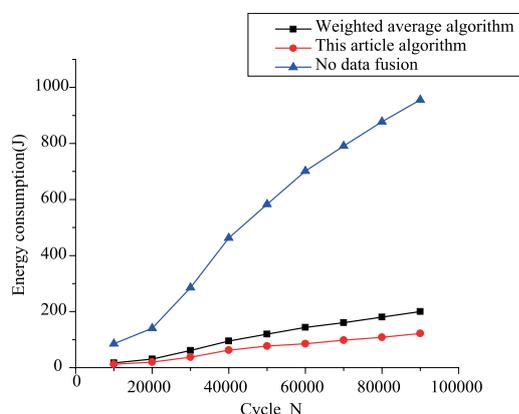


图 6 本文算法与未融合、加权均值算法总能耗比较

由图 6 仿真分析可知: 三种算法节点总能耗都随着循环周期的增加而增大, 其中本文两次阈值判定的数据融合算法能耗最小, 比加权平均算法能耗降低了 38.3%左右, 比没有进行任何融合算法降低了约 4 倍, 说明本算法能有效地节省节点能量。

#### 4 结束语

从图 3、4、6 的仿真分析可知: 本文用了一种使总能耗最小的分簇方法, 理论计算结果和仿真结果基本符合; 在进行簇头选择时, 考虑了节点剩余能量、簇头节点与簇内节点和 sink 之间的距离、节点当选为簇头的次数这些因素的影响, 相对 LEACH 中簇头的随机选择更能均衡节点的能耗; 在对采集的数据两次设置一定阈值减少网内数据的冗余, 比加权平均算法节省了 38.3%的能量, 同时能迅速辨别出异常瓦斯数据, 使工作人员更有效地对矿井进行监控, 对现代的煤矿

安全有很大的借鉴意义。

#### 参考文献

- 1 Chauhan S, Awasthi LK. Energy Efficient Data Gathering Protocol for Wireless Sensor Networks. Dept of CSE, National Institute of Technology, 2011: 535-535.
- 2 Zhao KQ, Wang F, Li MC. Wireless Sensor Network Data Aggregation Approach on Clonal Selection. Computer Technology and Application Progress, 2008:822-866.
- 3 Shen HY, Feng YM. Clustering Adaptive Weighted Fusion Algorithm of Wireless Sensor Network for Railway Embankment Monitor. Journal of Transportation Systems Engineering and Information Technology, 2010, 12(10): 190-194.
- 4 Kiran M, Kamal K, Gupta N. Study of Data Fusion in Wireless Sensor Network. Proc. of the International Conference on Advanced Computing and Communication Technologies (ACCT 2011): 535-539.
- 5 Mhatre V, Rosenberg C. Design Guidelines for Wireless Sensor Networks: Communication, Clustering and Aggregation. Ad Hoc Network Journal, 2004, 2(1): 45-63.
- 6 Heinzelman WR, Chandrakasan A, Balakrishnan H. An Application-Specific Protocol Architecture for Wireless Microsensor Networks. IEEE Transaction on Wireless Communications, 2002, 1(4): 660-670.
- 7 Zhou LP, Kamil M. Cluster Head Selection Algorithm in WSN Considering Security and Remaining Energy. CEA, 2012, 48(3): 88-89.
- 8 Zhao ZC, Deng SG, Jiang DC. Algorithm in Choice of Head-node on WSN based on Gaussian distribution. Computer Engineering and Applications, 2011, 47(16): 83-85.

(上接第 226 页)

- cuit interrupters. IEEE Trans. On Industry Applications, 2004, 40(4): 1006-1011.
- 7 窦甜华, 段培永, 段晨旭, 石嘉川. 一种建筑物低压供电线路故障电弧检测新方法. 微计算机信息, 2011, 27(3): 159-161.
  - 8 段培永, 窦甜华, 杨修文, 郭东东, 邹苒. 基于 CMAC 神经网络的低压故障电弧检测. 山东建筑大学学报, 2011, 26(2): 105-

109.

- 9 基于小波变换的单相接地故障电弧模型及其 PSCAD/EMTDC 仿真研究. 电力系统保护与控制, 2011, 39(5): 51-56.
- 10 尹同庆, 陈洪亮. 基于小波变换的故障电弧检测技术研究. 实验室研究与探索, 2008, 27(3): 19-21.
- 11 任龙霞, 吴为麟. 虚拟仪器在低压电弧故障断路器研制中的应用. 电力系统保护与控制, 2011, 39(5): 134-138, 143.