

基于模糊理论的无线传感器网络数据融合算法

窦桂琴¹, 万润泽², 张兴艳³

(1 中原工学院 计算机学院, 河南 郑州 450007; 2 湖北第二师范学院 计算机学院, 湖北 武汉 430205;

3 江汉大学 文理学院, 湖北 武汉 430056)

摘 要: 针对无线传感器数据融合过程中, 各节点观测值存在冗余以及监测区域重叠可能导致信息的精确性低、能量消耗大等问题, 本文提出利用模糊理论中的相关性函数计算节点间相互支持程度, 对支持程度高的传感器进行数据融合, 并利用融合结果与服务质量期望筛选出冗余节点, 使其进入休眠状态。仿真结果表明, 该方法能够获得更高的精度和可靠性, 并能有效延长网络生命周期。

关键词: 无线传感器; 数据融合; 置信距离

中图分类号: TP393

文献标识码: A

文章编号: 1000-7180(2012)09-0133-04

A Data Fusion Algorithm of Wireless Sensor Networks Based on Fuzzy Theory

DOU Gui-qin¹, WAN Run-ze², ZHANG Xing-yan³

(1 College of Computer, Zhongyuan University of Technology, Zhengzhou 450007, China;

2 College of Computer, Hubei University of Education, Wuhan 430205, China;

3 College of Arts and Science, Jiangnan University, Wuhan 430056, China)

Abstract: During the process of the data fusion in WSNs, the redundant highly correlated data and the overlapping region may lead to low precision of monitoring information and energy depletion. In this paper, we proposes the use of the correlation function in fuzzy theory to calculate the mutual support degree between the nodes, and compress the data from the nodes with higher mutual support degree. In order to reduce the energy consumption, we also consider that the redundant nodes, which should be changed into sleeping mode, could be chosen in terms of the fusion result and the expectation of QoS. The results demonstrate that the presented method can obtain much higher precision of collecting data and more reliable, also it can prolong the network lifetime effectively.

Key words: wireless sensor networks; data fusion; confidence distance

1 引言

随着无线传感器网络技术的发展, 如何实现传感器节点对网络中大规模异质数据的有效采集、减少存储空间的浪费以及降低信息转发功耗, 为目前函需解决的关键问题。数据融合是解决上述问题的基础, 其作用主要体现在节省能量、提高信息的精确性以及提高数据收集效率。

目前, 数据融合算法比较典型的有: 卡尔曼滤波、贝叶斯估计、D-S 论证推理、模糊逻辑、加权平均、统计决策理论等, 相对于融合理论, 数据融合算法更容易实现。文献[1]提出一种基于曲线拟合技术的流数据压缩方法, 对每个传感器采集的数据进行压缩, 最后在 Sink 节点进行还原。对于两个连续的采样时间, 即使节点采集的数据与前一时刻相同也进行转发, 因为没有考虑到数据的时间相关性, 从而

收稿日期: 2012-02-08; **修回日期:** 2012-03-10

基金项目: 湖北省教育厅项目(B20113007); 武汉市高校中青年教师项目(B2010033); 湖北省高校中青年教师创新团队(T200904); 湖北省教育厅科学技术研究重点项目(D2010302)

发送了冗余数据,造成了能量的浪费。

文献[2]提出通过时空域频率动态带宽的数据去时空冗余算法,对数据的时空冗余进行处理,降低网络通信开销。但是,传感器节点可能随着能量的下降,感知偏差或数据的不确定性将加大,依靠节点的位置和感知范围来对冗余节点进行决策可能降低网络的可靠性和有效性。

文献[3]通过概率方法筛选冗余节点,网络中分布的每个节点通过感知通信范围内的节点数量来计算自身可能成为冗余节点的概率,精确控制冗余节点。但是算法不足之处是没有考虑网络中的每个节点是如何通过与其他节点相互通信来统计半径内的邻居节点数量,以及这一部分开销。

本文根据模糊数学理论,提出了一种基于模糊理论的无线传感器数据融合算法,能够实现簇内保证簇头融合数据精度的条件下,让依存度较高节点适时进行休眠,减少簇内通信量,从而改善网络能量效率并延长网络生存周期。

2 基于模糊理论的数据融合算法

假设网络中有 n 个传感器, X_i 和 X_j 为第 i 和第 j 个传感器的感知数据,各节点测量的结果服从 Gauss 分布,概率分布函数为 $p_i(x)$ 和 $p_j(x)$, x_i 和 x_j 为其中一次观测值, x_i 服从正态分布 $N(\mu, \sigma_i)$, σ_i 为第 i 个传感器的测量精度。

定义 1 传感器 i 和传感器 j 之间测量数据的相互支持性,称为置信距离,可以采用 d_i 、 d_j 来表示,其值越小则说明两个传感器的测量值越接近,反之则表示两个传感器的测量值相差较大。

故节点 n_i 、 n_j 的感知数据置信距离测度 d_{ij} 为,

$$d_{ij} = 2 \int_{x_i}^{x_j} p_i(x | x_i) dx, d_{ji} = 2 \int_{x_j}^{x_i} p_j(x | x_j) dx, \text{ 其中}$$

$$p_i(x | x_i) = \frac{1}{\sqrt{2\pi}\sigma_i} e^{-\frac{(x-x_i)^2}{2\sigma_i^2}}, p_j(x | x_j) = \frac{1}{\sqrt{2\pi}\sigma_j} e^{-\frac{(x-x_j)^2}{2\sigma_j^2}},$$

$i, j = 1, 2, \dots, n$. 对网络中的 n 个传感器测量,置信距离 d_{ij} (其中 $i, j = 1, 2, \dots, n$) 可构成置信矩阵

$$D_n = \begin{bmatrix} d_{11} & d_{12} & \cdots & d_{1n} \\ d_{21} & d_{22} & \cdots & d_{2n} \\ \vdots & \vdots & & \vdots \\ d_{n1} & d_{n2} & \cdots & d_{nn} \end{bmatrix}, 0 \leq d_{ij} \leq 1, \text{ 且 } d_{ij} \text{ 越小}$$

说明节点 n_i 被 n_j 支持的程度越高。可令 $r_{ij} = 1 - d_{ij}$, 则由置信距离矩阵 D_n 可得到节点间关联矩阵 R

$$R = \begin{bmatrix} r_{11} & r_{12} & \cdots & r_{1n} \\ r_{21} & r_{22} & \cdots & r_{2n} \\ \vdots & \vdots & & \vdots \\ r_{n1} & r_{n2} & \cdots & r_{nn} \end{bmatrix}$$

其中, r_{ij} 表示节点 n_i 被节点 n_j 支持的程度,其值越大则支持程度越高。

一般的融合方法,是根据经验或多次测量的结果,给定一个阈值 β_{ij} [4], 令

$$r_{ij} = \begin{cases} 1, & d_{ij} \leq \beta_{ij} \\ 0, & d_{ij} > \beta_{ij} \end{cases}$$

这样,在关系矩阵 R 中存在三种关系:当 $r_{ij} = r_{ji} = 0$, 表示节点 n_i 被 n_j 相互不支持,相融性差; $r_{ij} = 1, r_{ji} = 0$, 表示节点 n_i 对 n_j 为弱支持; $r_{ij} = r_{ji} = 1$, 表示节点 n_i 被 n_j 相互支持,相融性好。

由关系矩阵 R 可以推算出各传感器测量的数据被其他传感器支持的程度,根据该重要程度可对其进行融合处理,以获得监测数据的较为精确的估计值。不过存在的问题是,阈值 β_{ij} 的选取过于主观,不利于对实际情况作出客观的辨别,阈值选取不当可能会对结果造成很大的影响。

考虑到只有当 $\sigma_i = \sigma_j$ 时,表示标准正态分布的取值,此时 $|x_i - x_j| / \sigma_i$ 才有明确的统计意义。而且,通常两个节点间的正向和反向数据支持程度一致,即 $d_{ij} = d_{ji}$, 为降低算法的复杂性,本文定义了一种复杂性较低的置信距离测算方法:

$$d_{ij} = d_{ji} = p(|Z| \leq \frac{|x_i - x_j|}{(\sigma_i + \sigma_j)/2})$$

同样,令 $r_{ij} = 1 - d_{ij}$, 得到节点间关联矩阵 R 。

引理 r_{ij} 是关于 $|x_i - x_j|$ 的下降函数,应具有较高的光滑性和单调性,确保在计算 r_i 时的稳定性。

根据信息融合一般理论, $x_1, x_2, \dots, x_n$ 的综合融合结果为: $x = (x_1 r_1 + x_2 r_2 + \dots + x_n r_n) / (r_1 + r_2 + \dots + r_n)$ 。

显然, $r_1, r_2, \dots, r_n$ 值的确定是数据融合的关键。对于计算得到的非负对称矩阵 R , 根据多元概率源信息合并理论 [5], 存在一组非负数 $\alpha_1, \alpha_2, \dots, \alpha_n$, 有

$$\begin{cases} r_i = \alpha_1 r_{i1} + \alpha_2 r_{i2} + \dots + \alpha_n r_{in} & i = 1, 2, \dots, n \\ \sum_{i=1}^n r_i = 1 \end{cases}$$

用矩阵形式表示上述线性方程组,为: $r = R\alpha$, 其中 $r = (r_1, r_2, \dots, r_n)^T$, $\alpha = (\alpha_1, \alpha_2, \dots, \alpha_n)^T$ 。根据非负对称矩阵的性质, R 存在最大特征值 λ , 满足 $R\alpha$

$=\lambda\alpha$, 有: $\lambda\alpha_l = \alpha_1 r_{l1} + \alpha_2 r_{l2} + \dots + \alpha_n r_{ln}$ 代入, 得到

$$x = \sum_{l=1}^n \frac{\alpha_l x_l}{\alpha_1 + \alpha_2 + \dots + \alpha_n}$$

3 簇内节点休眠调度

对于节点高密分布的网络, 如果节点感知数据跟融合结果之间偏差过大, 可以判断该节点可能受到环境的干扰, 其综合支持度降低, 但这并不能说明该节点彻底失效, 较好的处理方式是将这些节点看做冗余节点, 将其暂时设置为休眠状态。既节约这些节点的能量, 也可以降低对其他节点产生的干扰信号。簇内节点休眠调度算法的步骤如下:

(1) 簇头结点将簇内节点的测量值与融合结果进行比较, 计算 $\Delta x = \sqrt{(x - x_i)^2}$;

(2) 以 Δx 进行排序, 假设计算精度为小数点后两位, 考虑到一般冗余节点数量不会超过 60%, 为了降低计算复杂性, 采用基于非比较的基数排序算法, 算法复杂度为 $\Theta(n)$;

(3) 需确定可靠节点数的阈值。根据网络提出的服务质量期望^[6], 假设半径为 R 的圆形区域中的节点呈泊松分布, 设节点感知半径为 r , 则服务质量期望 Q 满足: $Q = 1 - e^{-\lambda\pi r^2}$ (λ 为满足期望的节点密度), 设满足服务质量期望 Q 的可靠节点数为 k , 由于 $\lambda = k/\pi r^2$, 故有 $k \geq R^2/r^2 \ln(1/1-Q)$ 。

(4) 根据可靠节点数量的最小值 k , 结合步骤(2)的排序结果, 筛选出要进入休眠状态的冗余节点, 它们将再间隔一个通信周期后被唤醒。

4 仿真实验和结果分析

为了评估算法的有效性, 以簇内节点数量为 10 予以验证。各传感器节点的测量值 $x_i = (0.95, 0.97, 1.20, 0.90, 1.22, 0.85, 1.14, 1.33, 0.93, 1.16)$, 标准正态分布取值为 $\sigma_i = (0.20, 0.15, 0.20, 0.30, 0.10, 0.25, 0.20, 0.05, 0.10, 0.30)$, 计算得 $x = \sum_{l=1}^n \frac{\alpha_l x_l}{\alpha_1 + \alpha_2 + \dots + \alpha_n} = 1.1346$ 。如果按照一般的融合方法, 给定阈值 $\beta_{ij} = 0.4, 0.6, 0.8$, 最后计算得 $x = 1.087, 1.120, 1.132$ 。显然, 在不同的阈值情况下得到的数据融合结果存在差异。而本文提出的算法通过将传感器之间的支持程度模糊化, 能有效避免事先给定阈值造成主观差异。

在文献[7]中, 提出过一种基于多跳的环切分层路由算法(Ring Layer Routing Algorithm)。算法的主要思想是在随机选择簇头后, 在簇内进行层次划

分。为验证一致性融合算法在能耗和传输效率方面的影响, 将前面提出的调度策略应用到该分簇方式中。通过 Java 语言编写的仿真程序来模拟该策略, 结合 Matlab 建立数学模型。实验参数: 网络覆盖范围 $(0, 0) \sim (200, 200)\text{m}$, Sink 位置 $(100, 100)$, 网络节点数 200, $E_{\text{elec}} = 50\text{nJ/bit}$, $E_{\text{amp}} = 100\text{PJ/bit/m}^2$, 数据报长度 $k = 1000\text{bits}$, 传感器初始能量 0.5J 。节点感知半径为 10m 。图 1 为应用了簇内节点休眠调度机制后, 与 RLRA 算法在网络生命周期上的比较, 可以发现头死亡节点的时间较 RLRA 明显延长, 但在 800 至 900 轮次的过程中, 分类后的节点死亡速度非常快, 其原因是采用模糊理论的数据融合算法中, 簇头计算其成员的综合支持度消耗较多能量, 各个簇头剩余能量较初始阶段能量大为减少, 因此综合支持度的计算加速了这些节点的死亡速度。

将簇内节点休眠调度策略中的服务质量期望的取值设定为 0.6、0.7、0.8 和 0.9, 传感器节点的感知半径分别设置为 5、10 和 15m, 头节点的死亡轮次对比情况如图 2 所示。可以看出, 在服务质量期望值

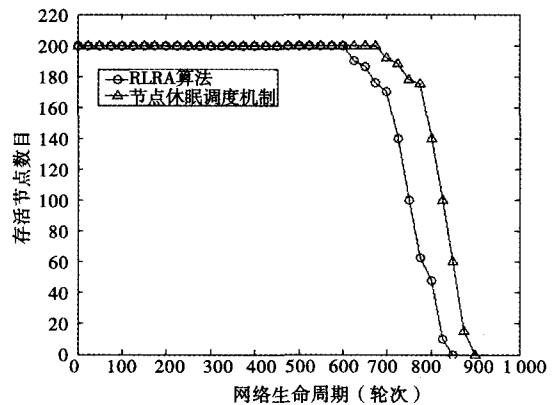


图1 网络生命周期的比较

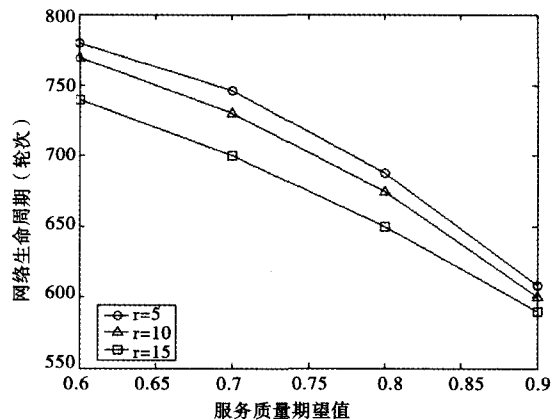


图2 不同服务质量期望下网络生命周期比较

取值较小时,节点感知半径越大则网络生命周期越长.因为节点感知半径越大,则成簇网络中的冗余节点数目越多,所需的最少可靠节点数则越少,这样可以降低簇内通信的能耗.

5 结束语

本文提出了一种一致性数据融合算法.算法定义了一种新的置信距离计算方法,并根据多元信息源合并理论对收集的测量值进行融合.为了保证一定服务质量的前提下提高网络生命周期,采用簇内节点休眠调度策略.通过仿真验证了算法的可行性,结果表明本算法能显著延长网络生命周期.

参考文献:

- [1] 刘向宇,王雅哲.面向无线传感器网络的流数据压缩技术[J].计算机科学与技术,2007,34(2):141-143.
- [2] 李国华,刘宝玲,沈树群.用于区域监测的无线传感器网络数据去冗余研究[J].微电子学与计算机,2005,22(9):134-136.
- [3] Gao Y, Wu K, Li F L. Analysis on the redundancy of wireless sensor networks [C]// Proceedings of the 2nd ACM International Conference on Wireless Sensor Networks and Applications. San Diego, USA: ACM, 2003: 108-114.
- [4] 涂国平,邓群钊.多传感器数据的统计融合方法[J].2001,20(3):28-32.
- [5] James O. Berger. Statistical Design Theory and Bayesian Analysis [M]. 北京:中国统计出版社,1998: 301-311.
- [6] Koskinen H. On the coverage of a random sensor network in a bounded domain[C]// Proceedings of the 16th International Tele-traffic Congress Specialist Seminar on Performance Evaluation of Wireless and Mobile Systems. Antwerp, Belgium: IAC, 2004: 11-18.
- [7] 侯西梅,万润泽.无线传感器网络的多跳分簇路由算法建模与实现[J].江西师范大学学报:自然科学版,2008,32(3):363-367.

作者简介:



钱桂琴 女,(1979—),讲师.研究方向为无线传感器网络.