

基于代理和熵权的无线传感器网络数据融合*

刘加杰, 孙子文

(江南大学 物联网工程学院, 江苏 无锡 214122)

摘要: 为了解决节点分布密集环境中网络数据融合能耗较大的问题,提出了一种基于代理和熵权的分组融合算法。该算法首先通过节点监测数据相似度的对比,剔除组内故障节点的影响;然后采用C/S模式对组内节点进行融合,以组融合数据代替节点数据参与融合;最后采用代理模式对各组融合数据进行熵权融合。仿真结果证明该方案能以较小的能耗代价获取准确的融合结果,有效减少了网络延迟。

关键词: 无线传感器网络; 数据融合; 代理; 熵权; 相似度

中图分类号: TP93

文献标志码: A

文章编号: 1001-3695(2012)10-3879-04

doi:10.3969/j.issn.1001-3695.2012.10.073

Agent and entropy-weight based WSN data fusion

LIU Jia-jie, SUN Zi-wen

(School of Internet of Things Engineering, Jiangnan University, Wuxi Jiangsu 214122, China)

Abstract: To overcome the huge energy consumption of data fusion in the sensor network cause by nodes dense deployment, this paper proposed an agent and entropy weight-based group data fusion algorithm. The algorithm first excluded the failure nodes' effect by the similarity comparison of nodes' monitoring data. And then obtained the fusion data of nodes by the entropy weight fusion of nodes with the model of C/S in groups, using the fusion datas obtained by C/S model to replace the monitoring datas of nodes when performed next fusion process. At last, fusioned the fusion dates with the model of agent between groups. The simulation results show that the proposed algorithm can obtain the accurate data fusion result with a smaller energy cost, and effectively in reducing network latency.

Key words: wireless sensor network (WSN); data fusion; agent; entropy weight; similarity

0 引言

无线传感器网络(WSN)由随机部署的节点组成,广泛应用于环境监测、战场监控等领域^[1,2]。它通过节点间的无线通信和连接,构成对某一特定环境或目标的监测网络,实现对环境或目标的监测。受节点自身能量和部署环境的影响,WSN技术的应用受到一定的限制。

数据融合技术能有效降低传感器节点的通信能耗,实现对目标的完整而准确地描述^[3,4]。文献[5]提出了一种可信融合算法,通过计算节点监测数据的可信度,选取可信度高的数据参与融合,提高融合结果的可信度。文献[6]提出了一种用于故障探测的基于模糊规则的数据融合方案,采用模糊集思想对节点监测数据进行融合。以上数据融合算法均基于客户端/服务器(client/serve, C/S)模式,节点将数据传送到汇聚节点或簇头节点集中进行处理,造成不必要的节点能耗且能耗不均^[7-9]。

文献[10]采用代理思想进行融合,通过代理在节点间的自由迁移将融合过程迁移到节点处进行,有效降低了节点的通信能耗。文献[11]提出了一种基于代理的多目标数据融合优化算法,由节点输出构造一个简单的重叠函数,通过函数变量

的求解判定目标存在的区域,实现对目标的定位和追踪。文献[12]将移动代理与神经网络相结合,通过节点监测数据的融合实现对目标的识别。

由上述可知,代理模式能有效降低节点通信能耗。但节点分布距离较近时存在两个问题:a)节点分布密集区采用代理模式,节点通信能耗较大;b)相邻节点的监测结果相近或相同,对各节点数据逐个进行融合会增大网络延迟。

针对问题a),本文将距离较近的节点分为一组,组内采用C/S模式,组间采用代理模式,以降低节点的通信能耗;针对问题b),本文通过组内节点监测结果的熵权^[13,14]融合,获取组融合数据,以组融合数据代替传感器节点监测数据参与融合,减小网络延迟。通过监测数据信息熵的计算,获取监测数据对应的熵权,通过加权融合,获取最终的数据融合结果。仿真结果表明,本文方案能获得准确的融合结果,并有效降低节点能耗和网络延迟时间。

1 熵权融合算法

熵的概念源于热力学,用于描述分子或离子运动的不可逆现象,后由香农引入信息论,用于对系统状态不确定性进行度量^[15]。设某一事件 x 发生的概率为 $p(x)$,则 x 的熵^[15]定义为

收稿日期: 2012-02-08; **修回日期:** 2012-03-10 **基金项目:** 国家自然科学基金资助项目(61174032);中央高校基本科研业务费专项基金资助项目(JUSR21131)

作者简介: 刘加杰(1987-),男,河南周口人,硕士研究生,主要研究方向为无线传感器网络(liujiajie20064300@163.com);孙子文(1968-),女,副教授,硕士,工学博士,主要研究方向为无线传感器网络、信息隐藏、模式识别、图像处理等。

$$H(x) = \log_2(p(x)) \quad (1)$$

其中: $H(x)$ 是对事件发生的不确定性的度量,其大小反映了事件发生的不确定性程度,其值越大,事件的不确定性越大。

所谓熵权就是根据各节点传递给决策者的数据信息熵的大小来决定节点数据的权重^[14]。它反映了不同节点的监测数据在决策中所起作用的大小。若某节点的监测数据在决策中所起的作用较大,则表示该节点数据所携带和传输的信息越多,它的信息熵就越小,对应的熵权就越大;反之,对应的熵权就越小^[14]。

设传感器节点 i 监测结果的概率向量为 $p_i = (p_i(A_1), p_i(A_2), p_i(A_3))$, p_i 中的元素 $P_i(A_j) (j=1,2,3)$ 分别表示节点监测结果中是目标的概率、不是目标的概率以及不确定性概率。

节点 i 的信息熵定义为

$$H_i = \sum_{j=1}^3 p_i(A_j) \log_2(p_i(A_j)) \quad (2)$$

节点 i 的监测结果的权重定义为

$$\partial_i = 1 - \frac{H_i}{\sum_{i=1}^n H_i} \quad (3)$$

节点 i 监测数据的熵权定义为

$$\omega_i = \frac{\partial_i}{\sum_{i=1}^n \partial_i} \quad (4)$$

满足归一化条件:

$$\sum_{i=1}^n \omega_i = 1 \quad (5)$$

其中: n 为参与融合传感器节点数目。

对满足相似度条件的网内各节点监测数据按式(6)~(8)进行融合,获取网内数据融合结果的概率向量为 $p = (p(A_1), p(A_2), p(A_3))$, 各分量分别为

$$p(A_1) = \sum_{i=1}^n \omega_i p_i(A_1) \quad (6)$$

$$p(A_2) = \sum_{i=1}^n \omega_i p_i(A_2) \quad (7)$$

$$p(A_3) = \sum_{i=1}^n \omega_i p_i(A_3) \quad (8)$$

数据融合结果的概率向量满足归一化条件的验证性推导:

$$\begin{aligned} p(A_1) + p(A_2) + p(A_3) &= \\ \sum_{i=1}^n \omega_i p_i(A_1) + \sum_{i=1}^n \omega_i p_i(A_2) + \sum_{i=1}^n \omega_i p_i(A_3) &= \\ \sum_{i=1}^n \omega_i (p_i(A_1) + p_i(A_2) + p_i(A_3)) &= \\ \sum_{i=1}^n \omega_i = 1 \end{aligned} \quad (9)$$

由式(9)可知,分组熵权融合算法融合结果的概率向量满足归一化条件,从而验证了分组熵权融合算法的正确性。

2 移动代理融合算法

本文融合方案中,融合代理分为节点融合代理(node fusion agent, NFA)和簇头融合代理(head fusion agent, HFA)两类。NFA 由簇头产生,负责簇内各组融合数据的融合,融合完成时,将融合结果返回给簇头节点;HFA 由汇聚节点生成,负责各簇的簇内融合结果的融合,簇间融合结束时,将融合结果

返回给汇聚节点。

本文基于代理的融合方案中,数据融合过程包括如下四个步骤:

a) 节点分组

对簇内节点进行分组,将距离较近的传感器节点分为一组。为各小组分配一个组长节点,由组长节点负责组内节点监测数据的融合。

b) 组内融合

组内各传感器节点将监测数据传送到组长节点。组长节点对监测数据进行累加求均值,计算各传感器节点监测数据与均值的相似度,并剔除相似度^[16]小的传感器节点的监测数据。采用熵权融合算法对满足相似度条件的传感器节点的监测数据进行融合,获取组融合数据,以组融合数据代替组内节点监测数据参与簇内融合。

假设组内传感器节点 i 的监测结果的概率向量为 P_i , 组内节点监测数据均值的概率向量 $\bar{R}_i$ 为

$$\bar{R}_i = (\bar{R}_i(A_1), \bar{R}_i(A_2), \bar{R}_i(A_3)) = \frac{1}{nn} \sum_{i=1}^{nn} p_i \quad (10)$$

其中: nn 为组内传感器节点的数目。

节点 i 与均值的距离定义为

$$d_i = \sqrt{\frac{1}{2} (\|p_i\| + \|\bar{R}_i\| + \langle p_i, \bar{R}_i \rangle)} \quad (11)$$

其中: $\|p_i\| = \langle p_i, p_i \rangle$ 。

内积:

$$\langle p_i, \bar{R}_i \rangle = p_i(A_1) \times \bar{R}_i(A_1) + p_i(A_2) \times \bar{R}_i(A_2) + p_i(A_3) \times \bar{R}_i(A_3)$$

节点 i 监测数据的相似度^[16] 定义为

$$\text{sim}(i) = 1 - d_i \quad (12)$$

设置相似度门限值,保留相似度大于门限值的节点监测数据,剔除相似度小于门限值的节点监测数据。根据式(1)~(4)获取满足相似度条件的各节点监测数据的熵权 ∂ ,经加权融合获得该组的组融合数据的概率向量 R_i 为

$$\begin{aligned} R_i = (R_i(A_1), R_i(A_2), R_i(A_3)) &= \sum_{i=1}^{nr} \partial_i \times p_i = \\ (\sum_{i=1}^{nr} \partial_i \times p_i(A_1), \sum_{i=1}^{nr} \partial_i \times p_i(A_2), \sum_{i=1}^{nr} \partial_i \times p_i(A_3)) \end{aligned} \quad (13)$$

其中: nr 为组内相似度大于门限值的节点数。

c) 簇内融合

簇头生成 NFA 并将其传送到第一跳组长节点,等待被响应。组长节点在响应 NFA 之前,通过对自身融合数据的概率向量 R_i 的运算和转换,获取熵权 ω 。NFA 被响应后,由 NFA 按式(14)对各组的组融合数据加权融合。融合完成后,NFA 携带融合结果迁移到下一跳组长节点继续进行加权融合。当簇内融合结束时,NFA 将簇内融合结果概率向量 H_i 返回给簇头节点,由簇头对融合结果进行下一步处理。

$$\begin{aligned} H_i = (H_i(A_1), H_i(A_2), H_i(A_3)) &= \sum_{i=1}^m \omega_i \times R_i = \\ (\sum_{i=1}^m \omega_i \times R_i(A_1), \sum_{i=1}^m \omega_i \times R_i(A_2), \sum_{i=1}^m \omega_i \times R_i(A_3)) \end{aligned} \quad (14)$$

其中: m 为簇 i 内分组的个数。

d) 簇间融合

汇聚节点生成 HFA 并传送至第一跳簇头节点,等待簇头节点的响应。簇头在响应 HFA 之前,对由 NFA 返回的簇内融

合结果进行分析,获取熵权 λ 。HFA 被响应后,由 HFA 按式 (15) 对簇内融合结果加权融合。融合完成后,HFA 携带融合结果迁移到下一跳簇头节点继续执行加权融合。簇间融合结束时,HFA 将簇间融合结果的概率向量 ST 返回给汇聚节点,由汇聚节点对融合结果进行最终的处理、分析和决策。

$$ST = (ST(A_1), ST(A_2), ST(A_3)) = \sum_{i=1}^{hn} \lambda_i \times H_i = (\sum_{i=1}^{hn} \lambda_i \times H_i(A_1), \sum_{i=1}^{hn} \lambda_i \times H_i(A_2), \sum_{i=1}^{hn} \lambda_i \times H_i(A_3)) \quad (15)$$

其中: hn 为簇头节点数。

3 节点能耗

C/S 融合模式中,节点能耗大致分为节点接收/发送数据的能耗、数据传输过程中的路径损耗、节点对数据进行融合的耗能三部分。与 C/S 融合模式相比,代理融合模式中节点能耗增加了节点接收/发送代理的耗能及传输代理过程中的路径损耗。

假设节点接收/发送数据的能耗系数为 α J/bit,融合能耗系数为 β J/bit,路径损耗系数为 γ J/bit/m;代理长度为 p bit,监测数据长度为 q bit。

节点 i 与 j 之间的距离定义为

$$d(i, j) = \sqrt{(x_i - x_j)^2 + (y_i - y_j)^2} \quad (16)$$

融合耗能定义为

$$E_f(q) = \beta \times q \quad (17)$$

C/S 模式中,节点接收/发送数据的能耗定义为

$$E_t(q) = E_r(q) = \alpha \times q \quad (18)$$

路径损耗定义为

$$E_l(q) = \gamma \times q \times d^4(i, j) \quad (19)$$

代理模式中,节点接收/发送数据的能耗定义为

$$E_t(p+q) = E_r(q) = \alpha \times (p+q) \quad (20)$$

路径损耗定义为

$$E_l(p+q) = \gamma \times (p+q) \times d^4(i, j) \quad (21)$$

节点完成一次融合总能耗定义为

$$E_s = E_t + E_r + E_f + E_l \quad (22)$$

4 仿真及结果分析

本文基于代理和熵权的无线传感器网络数据融合方案的 MATLAB 实验环境为:在 $100 \text{ m} \times 100 \text{ m}$ 的范围内随机生成 121 个节点的位置坐标,节点分布如图 1 所示;随机生成各节点监测结果的概率向量,作为数据融合实验中融合算法仿真的依据对本文算法进行仿真分析。

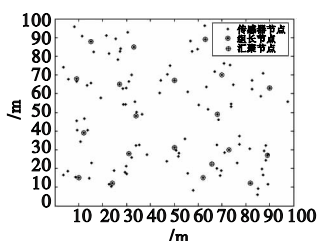


图1 节点分布

4.1 融合算法仿真

按照熵权融合算法分别对本文分组融合方案和原始不分组融合方案进行数据融合算法的仿真与分析。基于 MATLAB 平台对各数据融合阶段的实验仿真,获取本文分组融合方案的组内融合结果的概率向量如表 1 所示,本文分组融合模式和原始不分组融合模式的簇内融合结果的概率向量 CR 和簇间融合结果的概率向量 R 如表 2 所示。

表 1 组内融合结果

NO	$P(A_1)$	$P(A_2)$	$P(A_3)$	NO	$P(A_1)$	$P(A_2)$	$P(A_3)$
1	0.1611	0.1692	0.6697	11	0.1525	0.1706	0.6769
2	0.1648	0.1536	0.6716	12	0.1621	0.1436	0.6943
3	0.1575	0.1792	0.6633	13	0.1962	0.1364	0.6674
4	0.1668	0.1484	0.6848	14	0.2020	0.1140	0.6540
5	0.1735	0.1624	0.6641	15	0.1650	0.1554	0.6796
6	0.1606	0.1733	0.6661	16	0.1864	0.1488	0.6648
7	0.1757	0.1504	0.6739	17	0.1736	0.1399	0.6865
8	0.1806	0.1485	0.6709	18	0.1528	0.1836	0.6636
9	0.2064	0.1342	0.6594	19	0.1516	0.1516	0.6968
10	0.1663	0.1774	0.6563	20	0.1649	0.1647	0.6677

表 2 簇内/簇间融合结果

概率 向量	本文分组融合方案			原始不分组融合方案		
	$P(A_1)$	$P(A_2)$	$P(A_3)$	$P(A_1)$	$P(A_2)$	$P(A_3)$
CR	0.1626	0.1623	0.6751	0.1624	0.1634	0.6742
	0.1726	0.1586	0.6688	0.1724	0.1586	0.6690
	0.1714	0.1563	0.6723	0.1713	0.1562	0.6725
	0.1871	0.1463	0.6666	0.1888	0.1462	0.6650
	0.1607	0.1601	0.6792	0.1610	0.1593	0.6797
R	0.1708	0.1568	0.6724	0.1715	0.1562	0.6723

表 1 和 2 中的向量元素 $P(A_i)$ ($i=1,2,3$) 分别表示数据融合结果的概率向量中是目标的概率、不是目标的概率和不确定性概率。

表 1 给出了本文分组熵权融合算法中各组节点的组内融合数据的概率向量中的元素值。组融合数据是组内节点监测数据的总体特征性的代表,它反映了组内节点数据的决策趋向。按本文融合方案对表 1 中的组融合数据进行后续各阶段的融合,以获取相应的簇内融合结果及簇间融合结果。

表 2 显示了两组融合方案的簇内和簇间融合结果的概率向量。由表 2 中各项对应数据的对比可得:本文提出的分组熵权融合算法的融合结果与原始不分组融合算法的融合结果近似相同。由此可知本文提出的分组熵权融合算法能获取正确的融合结果,不会造成融合结果的失真,从而验证了本文分组熵权融合方案的正确性。

4.2 节点能耗仿真

假设网内传感器节点接收/发送数据的能耗系数 α 为 6×10^{-12} J/bit,数据融合的能耗系数 β 为 5×10^{-10} J/bit,路径损耗系数 γ 为 3×10^{-9} J/bit/m,代理长度 p 为 100 bit,监测数据长度 q 为 500 bit。按式 (13) ~ (19) 分别对代理模式及 C/S 模式下分组与不分组情况下的节点能耗进行仿真,能耗仿真如图 2 所示。

由图 2 可知,在节点分布相对较为集中的环境中,采用基于移动代理的分组融合模式能有效降低节点能耗。出现该情

况的原因是:a)节点能耗随着数据传输距离的增加而增加,一次性传输的距离越远,能耗越大,而分组融合模式能有效降低节点数据的一次性传输距离,从而降低节点能耗;b)代理融合模式通过代理的自由迁移将融合过程迁移到节点处进行,降低了节点的通信能耗。因此,在节点分布较为集中的环境中,将C/S模式与代理模式相结合,组内采用C/S模式,组间采用代理模式能有效降低节点的通信能耗。

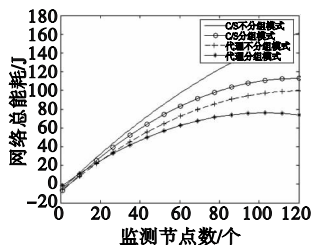


图2 节点能耗仿真

4.3 网络延迟仿真

网络延迟时间主要包括节点接收和发送数据产生的网络延迟、节点进行数据融合造成的网络延迟两个部分。而由节点进行数据融合造成的网络延迟只占很少的一部分。在 MATLAB 平台上分别对各融合模式下的网络延迟进行仿真,网络延迟仿真如图3所示。

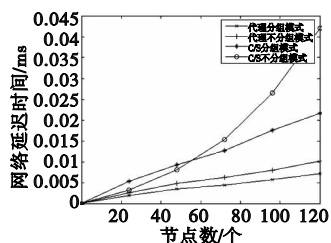


图3 网络延迟仿真

由图3可知,代理融合模式较C/S融合模式的网络延迟时间小;分组融合模式较不分组融合模式的网络延迟时间小。产生上述情况的原因是:a)代理模式使数据根据路由算法在簇内节点间进行传输,缩短了数据传输过程的延迟时间;b)与不分组融合模式相比,分组融合模式中,各组的组内融合过程同时进行能有效降低网络延迟时间。本文基于代理的分组融合算法通过分组融合与代理融合两种模式的有机结合,能有效降低数据融合过程中的网络延迟时间,提高数据的实时性。

5 结束语

本文通过分组模式、C/S模式与代理模式三种融合模式的结合,有效降低了在节点分布相对集中的环境中的节点能耗。通过对组内数据的相似度处理,剔除故障节点影响;通过组内融合获取组融合数据,以组融合数据代替组内节点的监测数据进行融合,以减小节点运算量、降低网络延迟;采用信息熵的思想,获取参与融合的各数据的熵权,通过节点数据的熵权融合获取融合结果。仿真结果证明了本文提出的数据融合算法能获得准确的融合结果,有效解决了节点分布相对集中环境代理融合模式能耗较大的问题,减少了网络延迟。

参考文献:

[1] ROGOVA G L, BOSSE E. Information quality in information fusion [C]//Proc of the 13th Conference on Information Fusion. 2010:1-8.

[2] POHANKA P, HRABOVSKY J, FIEDLER M. Sensor simulation environment for sensor data fusion [C]//Proc of the 14th International Conference on Information Fusion. Chicago:IEEE Press,2011:1-8.

[3] BAHREPOUR M, MERATNIA N, HAVINGA P J M. Sensor fusion based event detection in wireless sensor networks [C] //Proc of the 6th Annual International Mobile and Ubiquitous Systems: Networking & Services. Washington DC: IEEE Computer Society, 2009:1-8.

[4] XU Jian, YANG Geng, CHEN Zheng-yu, *et al.* Performance analysis of data aggregation algorithms in wireless sensor networks [C] //Proc of International Conference on Electrical and Control Engineering. Washington DC: IEEE Computer Society,2011:4619-4622.

[5] LUO Hong, TAO Hui-xiang, MA Hua-dong, *et al.* Data fusion with desired reliability in wireless sensor networks [C] //Proc of IEEE Transactions on Parallel and Distributed Systems. Washington DC: IEEE Computer Society, 2011:501-513.

[6] SHELL J, COUPLAND S, GOODYER E. Fuzzy data fusion for fault detection in wireless sensor networks [C]//Proc of Workshop on Computational Intelligence. Washington DC:IEEE Computer Society, 2010:1-6.

[7] WANG Li-qin, SHEN Lin. Research on the influence of data fusion on clustering lifespan in wireless sensor networks [C] //Proc of International Conference on Electrical and Control Engineering. Washington DC: IEEE Computer Society, 2010:2978-2982.

[8] LI Li, LI Wei-jia. The analysis of data fusion energy consumption in WSN [C] //Proc of International Conference on System Science, Engineering Design and Manufacturing Informatization. Washington DC: IEEE Computer Society,2011:310-313.

[9] SUNG W T, LIU Yu-feng, CHEN J H, *et al.* Enhance the efficient of WSN data fusion by neural networks training process [C] //Proc of International Symposium on Computer Communication Control and Automation. Washington DC: IEEE Computer Society, 2010: 373-376.

[10] ZHANG Li, WANG Qiang, SHU Xi-juan. A mobile agent based middleware for wireless sensor networks data fusion [C] //Proc of Instrumentation and Measurement Technology Conference. Washington DC: IEEE Computer Society, 2009:378-383.

[11] RAJAGOPALAN R. A multi-objective optimization approach for data fusion in mobile agent based distributed sensor networks [C]//Proc of IEEE Instrumentation and Measurement Technology Conference. Washington DC: IEEE Computer Society, 2010: 208-212.

[12] ZHOU Yan, WU Xia. Quantum neural network algorithm based on multi-agent in target fusion recognition system [C]//Proc of International Conference on Computational and Information Sciences. Washington DC: IEEE Computer Society,2010:537-540.

[13] MA Li-yi, XUE Wan-xin, GE Jian. Entropy method for decision-making of fuzzy information [C]//Proc of IEEE International Conference on Software Engineering and Service Sciences. Washington DC: IEEE Computer Society,2010:467-470.

[14] 万树平. 基于熵权的多传感器目标识别方法 [J]. 系统工程与电子技术,2009,31(3):500-502.

[15] 周荫清. 信息论基础 [M]. 北京:北京航空航天大学出版社,2008:14-31.

[16] 胡丽芳,关欣,何友. 基于可信度的证据融合方法 [J]. 信号处理, 2010,26(1):17-22.