# 无线传感器网络含有缺失数据的数据融合研究*

林 蔚, 尹 娟
(哈尔滨工程大学 理学院, 哈尔滨 150001)

**摘 要**: 针对含有缺失数据的无线传感器网络数据融合问题,提出了一种简单易行的二次数据融合算法(TFA)。考虑到感知数据的时—空相关性,对缺失数据进行线性插值插补和回归分析插补,对插补结果利用线性组合融合算法进行融合。综合考虑各节点的信息,利用自适应加权融合算法进行融合。实验结果表明,该算法在含有缺失数据的前提下,能够以较低的计算开销和较高的估计精度实现数据融合。

**关键词**: 无线传感器网络; 缺失数据; 数据融合; 线性组合融合; 自适应加权融合
**中图分类号**: TP393 **文献标志码**: A **文章编号**: 1001-3695(2012)11-4284-03
doi:10.3969/j.issn.1001-3695.2012.11.072

# Data fusion research based missing data in wireless sensor networks

LIN Wei, YIN Juan
(*College of Science*, *Harbin Engineering University*, *Harbin* 150001, *China*)

**Abstract**: This paper proposed a simple and feasible twice data fusion algorithm (TFA) to solve the problem of data fusion based data missing. As considered the sensing data's temporal and spatial correlation, it used the linear interpolation imputation algorithm and the regression analysis imputation algorithm to estimate the missing data, then used an adaptive linear combination fusion algorithm to estimate. As considered the data of each node, it also used an adaptive weighted fusion algorithm to estimate. The simulation results show that the algorithm has lower computing cost and better fusion effect with data missing case.

**Key words**: wireless sensor networks; data missing; data fusion; linear combination fusion; adaptive weighted fusion

## 0 引言

无线传感器网络(wireless sensor networks, WSN)的应用前景十分广阔,在军事、国防、工业、智能家居以及商务等许多重要领域都有着潜在的实用价值,被认为是对 21 世纪产生巨大影响力的技术之一[1]。WSN 综合了传感器、计算机网络、无线通信等多种先进技术,由部署在监测区域内大量的廉价微型传感器节点组成,通过无线通信形成一个多跳自组织网络系统,能够实时监测、感知和采集网络分布区域内监测对象的各种信息并加以处理,完成数据的采集和监测任务[2]。

由于传感器节点体积微小,电源能量、通信能力和计算能力有限,如何有效地去除冗余数据、减少网络通信量、提高资源的利用率,在节省能量的同时,形成比单一传感器更准确、更全面的融合值,是 WSN 数据融合研究的一个重要方面。数据的融合虽然会增加能耗,然而计算上的能耗与通信消耗的能量相比是 $10^{-3}$ 倍[3],因此,节能方面的优势可以得到充分体现。针对数据融合问题,很多学者已经做了一些研究工作,如模糊理论技术[4]、人工神经网络技术[5]、主成分分析技术[6]、贝叶斯估计技术[7]、D-S 证据理论技术[8]等均被应用到数据融合领域中,但这些方法都无法直接对含有缺失值的数据集进行融合。因此,在进行数据融合之前,对缺失的感知数据进行有效的插补是非常必要的。

考虑到传感器流数据的特点:a)数据之间有很强的关联性;b)数据之间具有明显的时—空特性,文献[9]研究了感知数据流上的缺失值插补问题,提出了 WARM 算法;文献[10]基于 WARM 算法,提出了其改进的 CARM 算法。然而如果在频繁模式中不存在缺失值所对应的数据元组,这两种算法均失效。为了有效解决上述算法所存在的问题,本文提出了一种线性组合融合算法,能够自适应地对缺失数据进行估计。文献[11]提出了基于感知数据时—空相关性的 HD 估计算法,但该算法的高精度是以计算时间为代价的,计算效率较低。

本文充分考虑了感知数据的时—空相关性,结合 LEACH 模型,提出了一种简单、易行的二次数据融合算法(TFA)。首先考虑到采集数据的时间相关性,利用线性插值插补算法对缺失数据进行估计。然后考虑到采集数据的空间相关性,利用回归分析插补算法对缺失数据进行估计。基于上述两种算法,通过利用缺失数据前后的真实数据,提出了一种基于感知数据时—空相关性的自适应线性组合融合算法。最后对线性组合融合算法处理后的不含缺失数据的各节点构造支持度矩阵,利用自适应加权数据融合算法进行二次融合。实验数据分析表明,该算法简单易行,融合效果较好。

**收稿日期**: 2012-04-02; **修回日期**: 2012-05-08 **基金项目**: 中央高校基础科研业务专项基金资助项目;黑龙江省博士后基金资助项目(3236301158)

**作者简介**: 林蔚(1964-),女,广东揭阳人,副教授,博士,主要研究方向为移动自组织网络、无线传感器网络等(linwei@hrbeu.edu.cn);尹娟(1986-),女,黑龙江佳木斯人,硕士.

## 1　线性组合融合算法

由于传感器节点电源能量有限且不可再生,当节点能量较低时,其工作状态不稳定,造成发送数据失败,从而导致数据缺失。并且,基于 WSN 的监测环境通常十分恶劣,可能使节点发生损坏或使通信覆盖范围只有几十到几百米的 WSN 的通信链路受到破坏而造成数据缺失。此外,还有一些其他因素,如节点时钟同步出现故障、路由算法选择失误等都有可能造成数据的丢失。基于以上原因,无论采用怎样有效的查询处理算法,在 WSN 中感知数据的缺失问题都是无法避免的。然而目前被广泛应用于数据融合领域的分析工具大多数无法处理含有缺失值的数据集,因此对缺失的感知数据进行有效的插补是非常必要的。

WSN 路由协议可以分为平面路由和分层路由两类。由于平面路由协议需要维持较大的路由表且占据的存储空间较多,因而并不适合在 WSN 中采用。分层路由算法可以在一定程度上解决这个问题,其中 LEACH(low energy adaptive clustering hierarchy,低功耗自适应分簇)算法是一种典型的分层路由算法,相邻的节点形成簇并选举簇头,成员节点将数据发送给簇头,由簇头融合数据并将结果发给汇聚节点。本文综合考虑了感知数据的时—空间相关性,结合 LEACH 模型,在簇头位置进行二次数据融合。图 1 为无线传感器网络分簇模型。

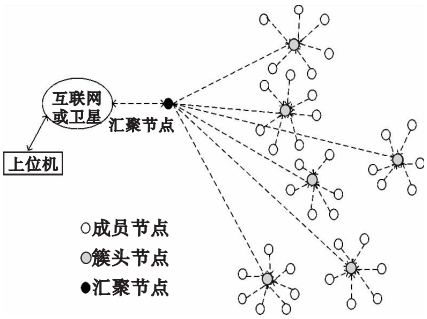


**图1　无线传感器网络分簇模型**

### 1.1　线性插值插补算法

WSN 监测的环境变量通常是连续变化的,即具有一定的时间相关性,可以充分利用这一特性对缺失的数据进行估计。由于线性插值函数能够较好地近似短时间内的函数变化曲线且易于计算,本文利用距缺失数据前后最近的非缺失数据构造线性插值函数来对缺失数据进行插补。

在簇头位置将 $n$ 个成员节点在采样周期 $T$ 内等时间间隔采集的 $\tau$ 次数据记为

$$X=\begin{pmatrix} x_1(1) & x_2(1) & \cdots & x_n(1) \\ x_1(2) & x_2(2) & \cdots & x_n(2) \\ \vdots & \vdots & & \vdots \\ x_1(\tau) & x_2(\tau) & \cdots & x_n(\tau) \end{pmatrix} \tag{1}$$

其中:每一行为不同成员节点在同一时刻采集的数据;每一列为同一成员节点在不同时刻采集的数据;$x_i(j)$ 为第 $i$ 个成员节点第 $j$ 次采集的数据,$i=1,2,\cdots,n$,$j=1,2,\cdots,\tau$。假设 $x_i(k)$ 缺失,首先找到与第 $k$ 次采集数据最近的两次非缺失数据 $x_i(k')$、$x_i(k'')$ $(k'\leqslant k\leqslant k'')$,根据式(2)进行线性插值插补。

$$\hat{x}_{i,L}(k)=x_i(k')+\frac{x_i(k'')-x_i(k')}{(k''-k')}(k-k') \tag{2}$$

其中:$\hat{x}_{i,L}(k)$ 为 $x_i(k)$ 线性插值插补的结果。但是当感知数据的时间相关性较低时,采用线性插值插补算法的估计误差会较大,此时可以利用回归分析插补算法来进行估计。

### 1.2　回归分析插补算法

由于地理位置相邻的传感器节点采集到的数据往往比较相似,即具有空间相关性,因此可以利用这一特性,对各节点采集到的数据利用 Pearson(皮尔森)相关系数方法进行空间相似度计算,找到与缺失数据节点空间相关性最大的传感器节点,对这两个节点的数据利用线性回归分析的方法进行插补。

当 $x_i(k)$ 缺失时,构造矩阵 $X'$:

$$X'=\begin{pmatrix} x_1(1) & \cdots & x_i(1) & \cdots & x_n(1) \\ \vdots & & \vdots & & \vdots \\ x_1(k-1) & \cdots & x_i(k-1) & \cdots & x_n(k-1) \\ x_1(k+1) & \cdots & x_i(k+1) & \cdots & x_n(k+1) \\ \vdots & & \vdots & & \vdots \\ x_1(\tau) & \cdots & x_i(\tau) & \cdots & x_n(\tau) \end{pmatrix}=\begin{bmatrix} X_1 & \cdots & X_i & \cdots & X_n \end{bmatrix} \tag{3}$$

为了找到与节点 $i$ 相关性最强的节点,将 $X'$ 中的每一列分别与第 $i$ 列求相关系数:

$$R(i,i')=\frac{\mathrm{cov}(X_i,X_{i'})}{\sqrt{D(X_i)}\sqrt{D(X_{i'})}}\quad i'=1,\cdots,i-1,i+1,\cdots,n \tag{4}$$

其中:$\mathrm{cov}(X_i,X_{i'})$ 为 $X_i$ 与 $X_{i'}$ 的协方差;$\sqrt{D(X_i)}$、$\sqrt{D(X_{i'})}$ 为 $X_i$ 与 $X_{i'}$ 的标准差。相关系数 $R(i,i')\in[-1,1]$,是 $X_i$ 与 $X_{i'}$ 之间线性关系强弱的一种描述性测量,或反映相关关系密切程度的重要指标。相关系数的绝对值越大,则表示这两个传感器采集的数据线性相关的程度越强。

利用式(5)找到与节点 $i$ 空间相关性最大的节点 $i''$:

$$R(i,i'')=\max|R(i,i')|\quad i''\in i' \tag{5}$$

构造节点 $i$ 和 $i''$ 的一元线性回归模型:

$$X_i=aX_{i''}+b+\varepsilon \tag{6}$$

其中:$a$ 和 $b$ 为回归系数,$\varepsilon$ 为随机误差项。采用最小二乘方法确定回归参数 $a$ 和 $b$ 后,得到缺失数据 $x_i(k)$ 的回归分析插补结果为

$$\hat{x}_{i,R}(k)=ax_{i''}(k)+b \tag{7}$$

### 1.3　线性组合融合算法

分析以上模型可知,当采集的数据出现缺失时,可以通过时间、空间相关性来对其进行插补。当数据间断性丢失、时间相关性较高时,可以利用线性插值插补算法对其进行估计;当数据连续性丢失时,线性插值插补算法效果较差,可以利用基于空间相关性的回归分析插补模型进行估计。然而在实际应用中,无法预知某段时间内数据的变化,所以无法确定使用哪种插补算法效果更好。本文综合考虑感知数据的时—空相关性,提出了一种线性组合融合算法。

为了能够自适应地求出线性插值插补和回归分析插补所对应的线性组合融合系数,利用缺失数据前后的非缺失数据求出线性组合融合系数,从而进行融合。

设 $\hat{x}_{i,L}(k-1)$、$\hat{x}_{i,L}(k+1)$ 为对非缺失数据 $x_i(k-1)$、$x_i(k+1)$ 利用线性插值插补的结果,$\hat{x}_{i,R}(k-1)$、$\hat{x}_{i,R}(k+1)$ 为对非缺失数据 $x_i(k-1)$、$x_i(k+1)$ 利用回归分析插补的结果。建立以下线性方程组:

$$\begin{cases} x_i(k-1)=\beta_L\hat{x}_{i,L}(k-1)+\beta_R\hat{x}_{i,R}(k-1) \\ x_i(k+1)=\beta_L\hat{x}_{i,L}(k+1)+\beta_R\hat{x}_{i,R}(k+1) \end{cases} \tag{8}$$

其中:$\beta_L$ 和 $\beta_R$ 为线性插值插补和回归分析插补的组合融合系数。可利用求解出的 $\beta_L$ 和 $\beta_R$,根据式(9)对缺失数据 $x_i(k)$ 进行线性组合融合。

$$\hat{x}_i(k) = \beta_L \hat{x}_{i,L}(k) + \beta_R \hat{x}_{i,R}(k) \tag{9}$$

## 2 自适应加权融合算法

由于无线传感器网络规模大且分布密度高,这样会引起监测区域的重叠,导致临近节点报告的信息存在一定程度的冗余性。为了能够将来自不同节点的测量数据进行综合分析和处理,在减少能量消耗的同时形成比单一传感器更可靠、更全面的融合值,需要对各节点的数据进行第二次融合。

将缺失数据线性组合融合的结果记录成矩阵的形式:

$$Y = \begin{pmatrix} y_1(1) & y_2(1) & \cdots & y_n(1) \\ y_1(2) & y_2(2) & \cdots & y_n(2) \\ \vdots & \vdots & & \vdots \\ y_1(\tau) & y_2(\tau) & \cdots & y_n(\tau) \end{pmatrix} \tag{10}$$

如果 $x_i(k)$ 为缺失数据,则 $y_i(k) = \hat{x}_i(k)$,否则 $y_i(k) = x_i(k)$。为了得到第 $k$ 次各节点采集数据的综合结果,本文对第 $k$ 行进行多传感器数据融合。如果 $y_i(k)$ 与第 $k$ 行其余数据的支持度越高,则 $y_i(k)$ 的真实性就越高;反之,真实性就越低。为了量化不同传感器在同一时刻观测值的支持程度,利用模糊数学中的隶属度函数来定义第 $k$ 次测量数据中传感器 $i$ 与 $j$ 之间的支持度为

$$s_{i,j}(k) = 2\operatorname{arccot}(|y_i(k) - y_j(k)|)/\pi \quad i,j = 1,2,\cdots,n \tag{11}$$

根据实际问题应用背景,可以设置一个支持度为 0 的距离阈值 $\delta \geqslant 0$;当 $|y_i(k) - y_j(k)| \geqslant \delta$ 时,$s_{i,j}(k) = 0$。则 $n$ 个传感器之间第 $k$ 次采集数据的支持度矩阵 $S(k)$ 为

$$S(k) = \begin{pmatrix} s_{1,1}(k) & s_{1,2}(k) & \cdots & s_{1,n}(k) \\ s_{2,1}(k) & s_{2,2}(k) & \cdots & s_{2,n}(k) \\ \vdots & \vdots & & \vdots \\ s_{n,1}(k) & s_{n,2}(k) & \cdots & s_{n,n}(k) \end{pmatrix} \tag{12}$$

$s_{i,j}(k)$ 仅表示测得数据 $y_j(k)$ 对 $y_i(k)$ 的支持程度,并不能反映系统中所有传感器所测得数据对 $y_i(k)$ 的支持程度,而需要综合 $S(k)$ 的第 $i$ 行全部元素,因而各传感器节点对应的加权系数 $w_i(k)$ 应综合各子系统 $s_{i,1}(k), s_{i,2}(k), \cdots, s_{i,n}(k)$ 的全部信息,所以需要求出一组非负数 $a_1(k), a_2(k), \cdots, a_n(k)$ 使

$$w_i(k) = a_1(k)s_{i,1} + a_2(k)s_{i,2} + \cdots + a_n(k)s_{i,n}$$

$$\sum_{i=1}^{n} w_i(k) = 1 \quad 0 \leqslant w_i(k) \leqslant 1 \tag{13}$$

假设 $W = [w_1(k), w_2(k), \cdots, w_n(k)]^{\mathrm{T}}$, $A = [a_1(k), a_2(k), \cdots, a_n(k)]^{\mathrm{T}}$,将式(13)改写成矩阵形式为

$$W = S(k)A \tag{14}$$

因为 $s_{i,j}(k) \geqslant 0$,所以矩阵 $S(k)$ 是一个非负矩阵。由非负对称阵的性质可知,$S(k)$ 存在最大模特征值 $\lambda(\lambda > 0)$ 及其对应的正特征向量 $A$,使得 $\lambda A = S(k)A$,则 $\lambda A$ 就可以作为多个节点测量数据的综合关联程度的度量,从而权重 $w_i(k)$ 即为 $A$ 归一化后的结果,因此

$$w_i(k) = \lambda a_i / \sum_{i=1}^{n} \lambda a_i = a_i / \sum_{i=1}^{n} a_i \tag{15}$$

则第 $k$ 次各节点采集的数据加权融合的结果为

$$y(k) = \sum_{i=1}^{n} w_i(k) y_i(k) \tag{16}$$

## 3 仿真实验

本文在 MATLAB 6.5 的环境下进行实验,以文献[12]中四个液压传感器节点对不同的压力状况进行实验,其测试值如表 1 所示。

表 1 四个传感器节点的八次测量数据

| 采集次数 | 传感器 1 | 传感器 2 | 传感器 3 | 传感器 4 | 真实值 |
|---|---|---|---|---|---|
| 第 1 次 | 4.93 | 4.98 | 4.94 | 4.97 | 5 |
| 第 2 次 | 4.94 | 5.01 | 4.94 | 5.01 | 5 |
| 第 3 次 | 7.96 | 7.88 | 7.98 | 7.86 | 8 |
| 第 4 次 | 9.96 | 9.84 | 9.96 | 9.75 | 10 |
| 第 5 次 | 11.96 | 11.94 | 11.96 | 11.61 | 12 |
| 第 6 次 | 17.98 | 17.92 | 17.99 | 17.91 | 18 |
| 第 7 次 | 22.93 | 22.76 | 21.02 | 22.75 | 22 |
| 第 8 次 | 25.03 | 24.87 | 25.03 | 24.88 | 25 |

为了测试实验效果,随机地将表 1 中节点 1 第 3 次采集的缺失数据和节点 4 第 5 次采集的缺失数据标记成缺失值 NaN1 和 NaN2,并对这些标记为缺失值的数据进行线性组合融合估计,通过用估计值与其对应的真实值进行比较来评价本文算法性能。

对 NaN1 利用其前后的真实数据得到线性插值插补系数和回归分析插补系数分别为 -0.042 3 和 1.053 7,对 NaN2 利用其前后的真实数据得到线性插值插补系数和回归分析插补系数分别为 -0.068 3 和 1.069 7,从而采用本文方法进行第 1 次线性组合融合的结果为 NaN1 = 7.9887, NaN2 = 11.8053。

为了评价算法的性能,本文利用线性组合融合后的结果与原始值的均方根误差 RMSE 作为度量标准,并与线性插值插补算法、回归分析插补算法、平均值插补算法进行了比较,如表 2 所示。

$$\mathrm{RMSE} = \sqrt{\frac{\sum_{i=1}^{n} (\hat{x}_i(k) - x_i(k))^2}{n}}$$

其中:$n$ 为缺失数据的个数。

表 2 第 1 次融合结果及各种方法估计误差

| 估计结果 | 线性插值 | 回归分析 | 平均值 | 线性组合融合 | 真实值 |
|---|---|---|---|---|---|
| NaN1 | 7.450 0 | 7.880 5 | 7.906 7 | 7.988 7 | 7.96 |
| NaN2 | 13.830 0 | 11.918 9 | 11.953 3 | 11.805 3 | 11.61 |
| RMSE(Pa) | 1.610 7 | 0.225 5 | 0.245 7 | 0.139 6 | |

从表 2 中可以看出,综合考察感知数据的时空相关性,使用本文线性组合融合后的效果较单独基于时间相关性的线性插值算法和单独基于空间相关性的回归分析算法要好,且比使用平均值算法要好。

对利用线性组合融合后的数据进行自适应加权融合,以第 1 次各传感器的测量值为例,得到最大模特征值为 3.942 7,对应的特征向量为 0.499 2、0.499 2、0.500 8、0.500 8,从而得到第 1 次测量各传感器的权 $w = [0.2496, 0.2496, 0.2504, 0.2504]$,融合结果为 4.955 0,误差为 0.045。同理,可得其他次融合结果及误差,如表 3 所示。

表 3 第 2 次融合结果及误差

| 采集次数 | 融合结果 | 误差 | 采集次数 | 融合结果 | 误差 |
|---|---|---|---|---|---|
| 第 1 次 | 4.955 0 | 0.045 0 | 第 5 次 | 11.917 6 | 0.082 4 |
| 第 2 次 | 4.975 0 | 0.025 0 | 第 6 次 | 17.950 0 | 0.050 0 |
| 第 3 次 | 7.927 2 | 0.072 8 | 第 7 次 | 22.562 4 | 0.562 4 |
| 第 4 次 | 9.878 5 | 0.121 5 | 第 8 次 | 24.952 6 | 0.047 4 |

从表 3 中可以看出,对第 3 次和第 5 次含有 (下转第 4308 页)

转换直接等价于 $\Omega$ 下的评估数据,其他识别框架上的评估数据需根据规则转换矩阵转化到 $\Omega$ 上,其中规则转换矩阵分别为

$$\begin{pmatrix} 1 & 0 & 0 & 0 & 0 \\ 0.3 & 0.7 & 0 & 0 & 0 \\ 0 & 0.4 & 0.6 & 0 & 0 \\ 0 & 0 & 0.4 & 0.6 & 0 \\ 0 & 0 & 0 & 0.4 & 0.6 \\ 0 & 0 & 0 & 0 & 1 \end{pmatrix}_{6\times5} \begin{pmatrix} 0.8 & 0.2 & 0 & 0 & 0 \\ 0 & 0.4 & 0.6 & 0 & 0 \\ 0 & 0 & 0.6 & 0 & 0 \\ 0 & 0 & 0.4 & 0.6 & 0 \\ 0 & 0 & 0 & 0.2 & 0.8 \end{pmatrix}_{4\times5} \begin{pmatrix} 0.8 & 0.2 & 0 & 0 & 0 \\ 0 & 0.2 & 0.6 & 0.2 & 0 \\ 0 & 0 & 0 & 0.2 & 0.8 \end{pmatrix}_{3\times5}$$

运用效用准则划分区间 $[0,1]=[0,0.2]\cup[0.2,0.45]\cup[0.45,0.6]\cup[0.6,0.8]\cup[0.8,1]$,分别对应 $\Omega$ 的五个等级。按照前文叙述的步骤和方法逐步进行转换、合成,由下至上逐级合成评估模型中的评估数据得 $B(V)=\{(H_1,0.0785),(H_2,0.2024),(H_3,0.2818),(H_4,0.4454),(H_5,0.1264)\}$,$u(B(V))=0.71$。

根据效用准则可得评估等级对应 $\Omega$ 中的"良好",因此建议采用云计算模式解决研究所的当前问题。评估结果没有呈现出显著优势的一个重要原因在于用户对云计算服务模式在安全和服务透明度方面的担忧,这也与现实情况相符合。

## 4 结束语

尽管云计算目前在商业界取得了一定的应用,但其目前的技术和管理水平离开放、互通、融合、安全和标准都尚有一段距离,这些问题都会对云计算的广泛应用产生重大影响,因此用户需要对云计算的综合能力进行有效评估才能作出正确的决策。本文首先构建了一套云计算服务适应性评估指标体系,而后运用证据推理的方法选择一个算例进行验证,并取得了比较合理的结果,进一步证明了本文构建的评估指标体系的有效性和合理性。

下一步将进一步研究针对不同应用场景下的服务适应性以及服务提供商选择的问题。

**参考文献:**


[1] MARSTON S, LI Zhi, BANDYOPADHYAY S, *et al*. Cloud computing: the business perspective [J]. Decision Support Systems, 2011, 51(1): 176-189.

[2] ARMBRUST M, FOX A, GRIFFITH R, *et al*. Above the clouds: a berkeley view of cloud computing [EB/OL]. [2007-02-10]. http://www.eecs.berkeley.edu/Pubs/TechRpts/2009/EECS-2009-28.html.

[3] 迟学斌. 云计算不适合科学计算[EB/OL]. [2008-12-11]. http://server.chinabyte.com/139/8620139.shtml.

[4] "云计算"行业热点问题分析——五大问题[EB/OL]. [2011-01-12]. http://vccbook.blog.51cto.com/1589501/478151.

[5] 蔡盈芳. 基于云计算的信息系统安全风险评估模型 [J]. 中国管理信息化, 2010, 13(12): 75-77.

[6] MISRA S C, MONDAL A. Identification of a company's suitability for the adoption of cloud co9mputing and modelling its corresponding Return on Investment [J]. Mathematical and Computer Modelling, 2011, 53(3-4): 504-521.

[7] KRUGMAN P R, OBSTFELD M. International economics: theory and policy, economies of scale [M]. 6th ed. [S. l.]: Imperfect Competition & International Trade, 2003.

[8] 罗军舟,金嘉晖,宋爱波,等. 云计算:体系架构与关键技术[J]. 通信学报, 2011, 32(7): 3-20.

[9] DARGHA R. Cloud computing, infosys white paper[S]. 2009.

[10] 里特豪斯. 云计算实现、管理与安全 [M]. 田思源,赵学锋,译. 北京:机械工业出版社, 2010: 105-123.

[11] 冯登国,张敏,张妍,等. 云计算安全研究[J]. 软件学报, 2011, 22(1): 71-83.

[12] Distributed Management Task Force. DSP-IS0101, Interoperable clouds, a white paper from the open cloud standards incubator[S]. 2009.

[13] 马泽尔. 云计算安全与隐私 [M]. 刘戈舟, 杨译明,刘宝旭,译. 北京:机械工业出版社, 2011: 63-66.

[14] YANG Jian-bo. Rule and utility based evidential reasoning approach for multiattribute decision analysis under uncertainties [J]. Euripean Journal of Operational Research, 2001, 131(1): 31-61.

[15] WANG Ying-ming, YANG Jian-bo, XU Dong-ling. Enciromental impact assessment using the evidence reasoning approach [J]. European Journal of Operational Research, 2006, 174(3): 1885-1913.


(上接第 4286 页)缺失数据的各节点经过第 1 次线性组合融合后,再利用第 2 次自适应加权融合的误差较小。

## 4 结束语

针对含有缺失数据的无线传感器网络数据融合问题,提出了一种简单易行的二次数据融合算法(TFA)。本文充分考虑无线传感器网络的时—空特性,对缺失数据利用线性组合融合方法进行第 1 次数据融合,融合效果要比只考虑时间相关性的线性插值方法和只考虑空间相关性的回归分析方法效果好。为了能够形成比单一传感器更准确、更全面的观测值,对各传感器同次采集的数据进行第 2 次自适应加权融合,融合效果较好。

**参考文献:**


[1] 李建中,李金宝,石胜飞. 传感器网络及其数据管理的概念、问题与进展[J]. 软件学报, 2003, 14(10): 1717-1727.

[2] 崔莉,鞠海玲,苗勇,等. 无线传感器网络研究进展[J]. 计算机研究与发展, 2005, 42(1): 163-174.

[3] 张强,杨涛. 用于环境监测的自供电传感器网络[J]. 仪表技术与传感器, 2008(2): 34-36.

[4] 王飞. 无线传感器网络中基于模糊理论的决策级数据融合技术研究[D]. 北京:北京交通大学, 2010.

[5] 孙凌逸,黄先祥,蔡伟,等. 基于神经网络的无线传感器网络数据融合算法[J]. 传感技术学报, 2011, 24(1): 122-127.

[6] 马守明,王汝传,叶宁. 基于主元分析的无线传感器网络数据融合研究[J]. 计算机工程与科学, 2010, 32(11): 44-46, 65.

[7] 张书奎,崔志明,龚声蓉,等. 基于 Bayes 序贯估计的无线传感器网络数据融合算法[J]. 电子与信息学报, 2009, 31(3): 716-721.

[8] 任秀丽,田洋. 传感器网络中基于预处理证据理论的数据融合[J]. 计算机应用, 2011, 31(7): 1992-1994.

[9] HALATCHEV M, GRUENWALD L. Estimating missing values in related sensor data streams[C]//Proc of the 11th International Conference on Management of Data. 2005: 83-94.

[10] JIANG Nan, GRUENWALD L. Estimating missing data in data streams [C]//Proc of the 12th International Conference on Database Systems for Advanced Applications. 2007: 981-987.

[11] 何丽娟,周鸣争,陶皖,等. 无线传感器网络中不确定数据的估计算法[J]. 计算机工程与应用, 2011, 47(28): 100-102, 161.

[12] 蔡科,左宪章,康健,等. 无线传感器网络测试系统数据融合研究[J]. 计算机仿真, 2010, 27(7): 1-4, 9.