# 无线传感器网络中基于序列相关性的数据压缩算法

翟　双[①②]　　钱志鸿*[①]　　刘晓慧[①]　　孙大洋[①]
[①](吉林大学通信工程学院　长春　130012)
[②](长春工业大学计算机科学与工程学院　长春　130012)

**摘　要**：无线传感器网络(WSN)中传输的数据具有相关性和冗余性。如何有效降低网络中的数据量，延长网络生命周期，始终是 WSN 的研究热点之一。该文基于 WSN 中数据序列的相关性，提出一种两步数据压缩算法(TSC-SC)。网络中的簇首和簇内节点执行各自的压缩算法：簇首首先执行相关性分组算法，将数据分组，减少簇内节点的计算量以及消除簇内数据的空间相关性；簇内节点对多属性数据分类压缩，并将压缩参数传至簇首，簇首解压后再次进行分类压缩，进一步消除数据相关性，减少节点数据冗余度，降低通信能耗。为实现对压缩算法的综合性能评价，考虑基本的压缩要求和算法的计算能耗，提出了基于能量判别的算法评估模型(NCER)。仿真结果表明 TSC-SC 算法可以有效降低压缩比和压缩误差，充分减少数据传输量和网络的通信能耗，利用 NCER 指标能够直观地评价算法的性能。

**关键词**：无线传感器网络；数据压缩；序列相关性
**中图分类号**：TP393　　**文献标识码**：A　　**文章编号**：1009-5896(2016)03-0713-07
**DOI**: 10.11999/JEIT150280

# Data Compression Algorithm Based on Sequence Correlation for WSN

ZHAI Shuang[①②]　　QIAN Zhihong[①]　　LIU Xiaohui[①]　　SUN Dayang[①]
[①](*College of Communication Engineering*, *University of Jilin*, *Changchun* 130012, *China*)
[②](*Institute of Computer Science and Engineering*, *Changchun University of Technology*, *Changchun* 130012, *China*)

**Abstract**: The data has correlations and redundancy in Wireless Sensor Network (WSN). How to reduce effectively the amount of communication data and extend the network life cycle is one of researching hot points. The Two-Step data Compression algorithm based on Sequence Correlation (TSC-SC) for WSN is proposed in this paper. The cluster head and the nodes in clusters perform different compression algorithms for themselves. In order to eliminate the spatial correlation of data and reduce the calculated amount, the cluster head nodes perform the grouping algorithm firstly, then the nodes in clusters perform the classifing compression to eliminate correlation for multi-attribute data, and pass the compression parameters to the cluster head; the cluster head perform the classifing compression again after decompressing the parameters. So the data-redundancy and communication energy consumption is further reduced. A new evaluation model named Network Compression Energy Ratio (NCER) based on energy discrimination is also proposed. The evaluation model realizes comprehensive evaluation of compression algorithms by considering both the basic requirements of compression and calculated energy consumption in the nodes. Simulation results show that TSC-SC algorithm can reduce the compression ratio and compression error effectively; the amount of communication data and energy consumption can achieve a satisfactory level in the network. The algorithm can be estimated directly using NCER.

**Key words**: Wireless Sensor Network (WSN); Data compression; Sequence correlation

## 1　引言

无线传感器网络(Wireless Sensor Network, WSN)的能量问题一直是限制其广泛应用的瓶颈，也是研究中要解决的关键问题[1,2]。无线传感器网络作为新一代的通信网络，具有能力受限、节点分布密集的特点。WSN 中有很多的关键技术值得关注，数据压缩技术是 WSN 中的关键技术之一。采用数据压缩可以有效减少网络中的数据量，降低通信能耗，延长网络生命周期。因此，数据压缩技术已经

收稿日期：2015-03-09；改回日期：2016-01-06；网络出版：2016-01-26
*通信作者：钱志鸿　dr.qzh@163.com
基金项目：国家自然科学基金(61371092)
Foundation Item: The National Natural Science Foundation of China (61371092)

成为了 WSN 的研究热点之一[3,4]。现有数据压缩算法的核心思想均为如何去除数据相关性，降低冗余度，达到减少网络中数据传输量的目的，在 WSN 中实现低能耗、高质量的数据传输和处理。但由于无线传感器网络具有应用相关的特点，不同压缩算法所针对的数据冗余种类不同，因此，在一定程度上也制约了算法的压缩效果，限制了算法的应用范围。

WSN 中的数据通常可将其称为多源信息[5,6]，这些数据的相关性存在于很多方面：一个传感器节点(Sensor Node, SN)在一段时间内采集到的数据会有一定的冗余，这就是数据的时间相关性；通常为了保证网络数据传输的可靠性，SN 的部署一般都比较密集，相邻 SN 的监测区域会产生重叠，在重叠区域监测的数据必定存在相似性，即数据的空间相关性；随着人们对网络服务与应用需求的日益增长，网络提供的服务种类越来越多，由单个 SN 对覆盖区域内的多属性的数据同时进行采集和控制已经成为了可能，多属性数据之间的相关性即为属性相关性。

本文将 WSN 中原始数据存在的这些相关性称为数据的多维相关性。同时考虑 WSN 中多维数据相关性的压缩算法国内外并不多见，大多数只是针对空时相关性进行研究。例如文献[7]中提出的分布式结构树压缩算法，利用节点数据中存在的时间、空间相关性，在基站处根据预先求得的相关系数及少量原始数据恢复全部原始数据。文献[8]提出的基于回归模型的 CODST 算法，利用曲线拟合技术实现对空时数据压缩。文献[9]提出了一种自适应最优消零压缩算法，利用数据的时空相关性去除冗余。目前，也有很多学者利用压缩感知理论，实现数据压缩，如文献[10]利用了压缩感知理论实现了对 2 维几何信号的压缩。文献[11]基于压缩感知理论，设计了一种具有长生命周期的数据收集方法，实现数据压缩。文献[12]提出基于数据融合树的压缩算法，通过减小投影矩阵与稀疏基之间的相关度重构数据，达到数据压缩目的。文献[13]通过匹配 DTW 路径中的数据与异步数据之间的函数关系，得到数据矩阵，再利用数据之间的相关性进行数据压缩。但是上述算法都仅消除了数据的时间与空间相关性，并没有对属性相关性进行考虑。文献[14]中提出的基于小波的自适应多模数据压缩算法，尽管考虑了多维相关性，但是都没考虑 SN 的数据分组、能量消耗和节点的存储能力等影响。

因此，考虑数据的多维相关性，本文提出一种基于多维序列相关性的数据压缩算法，并设计了基于能量判别的算法评估模型。通过仿真实验证明，该算法能有效减少网络中的数据量，延长网络生命周期，同时又可达到较好的节能效果。

## 2 预备知识

### 2.1 数据存储格式

**2.1.1 原始采集数据的存储格式** 设每个SN可采集到$N$种数据，分别用$\boldsymbol{X}_1, \boldsymbol{X}_2, \cdots, \boldsymbol{X}_I, \cdots, \boldsymbol{X}_N$表示。每种属性的数据都是一个长为$M$的时间序列，即$\boldsymbol{X}_I = (x_{i1}, x_{i2}, \cdots, x_{ii}, \cdots, x_{im})^{\mathrm{T}}$，其中，$x_{ij}$表示$j$时刻上，第$i$种属性数据的监测值。则该SN的原始采集数据可用矩阵$\boldsymbol{S}$来表示，即

$$\boldsymbol{S} = \begin{bmatrix} \boldsymbol{X}_1 & \boldsymbol{X}_2 & \cdots & \boldsymbol{X}_N \end{bmatrix} = \begin{bmatrix} x_{11} & x_{21} & \cdots & x_{n1} \\ x_{12} & x_{22} & \cdots & x_{n2} \\ \vdots & \vdots & \ddots & \vdots \\ x_{1m} & x_{2m} & \cdots & x_{nm} \end{bmatrix} \tag{1}$$

**2.1.2 压缩数据的存储格式**

(1)$\mathrm{Data}_{\mathrm{Sa_I}} = (b_I, k_I)$为压缩参数，表示属性为 $I$ 的所有相关序列压缩后的存储格式，$b_I$ 与 $k_I$ 分别为该序列的线性拟合系数。

(2)$\mathrm{Data}_{\mathrm{Sa_I}} = (T_{\mathrm{st_I}}, T_{\mathrm{en_I}}, \mathrm{CoefVec_I})$为基序列与独立序列压缩后的存储格式，表示属性为 $I$ 的基序列与独立序列在时间段$(T_{\mathrm{st_I}}, T_{\mathrm{en_I}})$的压缩参数。其中，$T_{\mathrm{st_I}}$ 与 $T_{\mathrm{en_I}}$ 分别表示数据序列在当前时间段的起点与终点，$\mathrm{CoefVec_I}$表示拟合多项式系数，即$\mathrm{CoefVec_I} = (a_0, a_1, \cdots, a_n)$。

当节点对压缩后的数据进行传输时，再加上节点自身的 ID 号即可。

### 2.2 相关性描述

本文参考文献[14]中的数据相关度来描述两个序列 $\boldsymbol{X}$ 与 $\boldsymbol{Y}$ 之间的线性相关程度。为了书写方便，本文中用 $\rho_{\boldsymbol{XY}}$ 来表示相关系数，具体公式及意义在此不再赘述。

**定义 1** 对于$\forall \boldsymbol{X}_I, \boldsymbol{X}_J \in \boldsymbol{S}$，以及给定的相关性阈值 RelThreshold，若 $\boldsymbol{X}_I$ 与 $\boldsymbol{X}_J$ 的相关系数 $\rho_{IJ} >$ RelThreshold，则称 $\boldsymbol{X}_I$与 $\boldsymbol{X}_J$相关；反之，若 $\rho_{IJ} \le$ RelThreshold，则称 $\boldsymbol{X}_I$与 $\boldsymbol{X}_J$不相关。

$$\left.\begin{array}{ll} \boldsymbol{X}_I\text{与}\boldsymbol{X}_J\text{相关}, & \rho_{IJ} > \mathrm{RelThreshold} \\ \boldsymbol{X}_I\text{与}\boldsymbol{X}_J\text{不相关}, & \rho_{IJ} \le \mathrm{RelThreshold} \end{array}\right\} \tag{2}$$

**定义 2** 对于$\forall \boldsymbol{X}_I \in \boldsymbol{S}$，矩阵 $\boldsymbol{S}$ 中，与 $\boldsymbol{X}_I$ 相关的序列个数定义为 $\boldsymbol{X}_I$ 的相关次数$\mathrm{RelNum}_I$。

$$\mathrm{RelNum}_I = \sum_{J=1, J\ne I}^{N} \left[ \rho_{IJ} - \mathrm{RelThreshold} \right] \tag{3}$$

**定义 3** 对于$\forall \boldsymbol{X}_I \in \mathrm{CanBase}$，定义相关总和$\mathrm{RelSum}_I$为矩阵 $\boldsymbol{S}$ 中与 $\boldsymbol{X}_I$相关的所有相关系数之和

(此处 CanBase 为候选基集合)。

$$\mathrm{RelSum}_I = \sum_{J=1,J\neq I}^{N} \rho_{IJ} \tag{4}$$

## 3　基于序列相关性的数据压缩算法

本文提出的 WSN 中基于序列相关性的数据压缩算法(Two-Step data Compression algorithm based on Sequence Correlation, TSC-SC)分为两个部分。首先，簇首执行相关性分组算法，将监测到的数据分组，减少簇内节点的计算量，同时簇首将分组信息广播给簇内节点；簇内节点接收到簇首的广播信息后，对本地节点数据执行分类压缩，消除数据之间的时间以及属性相关性，再将压缩参数发送至簇首，簇首根据接收到的压缩参数，对数据再次执行分类压缩，并将去除了相关性的数据传送给 Sink 节点。

### 3.1 数据的相关性分组

算法首先对多属性数据进行相关性分组，选出候选基序列并由此确定基序列，在此基础上，选出独立序列以及相关序列。其中，基序列的个数影响算法的压缩比，基序列与其相关序列的相关程度对算法的压缩精度会产生影响，因此基序列的确定是分组的关键。文中针对簇首提出了一种数据相关性分组算法，通过对基序列的筛选，最大程度地实现对相关序列的精确表示，在原有算法的基础上进一步提高压缩效率。

#### 3.1.1 选取基序列

(1)选择候选基序列：　分别计算 $\boldsymbol{S}$ 中所有序列的相关次数 $\mathrm{RelNum}_I$，若某序列 $\boldsymbol{X}_I$ 的相关次数为最大值且不为 0，则 $\boldsymbol{X}_I$ 被选为候选基序列，放入候选基集合 CanBase 中。

$$\mathrm{CanBase} = \left\{\boldsymbol{X}_I \middle| \mathrm{RelNum}_I = \max, \text{且}\, \mathrm{RelNum}_I \neq 0,\ 0 < I \leq N\right\} \tag{5}$$

(2)选择基序列：　对于 $\forall \boldsymbol{X}_I \in \mathrm{CanBase}$，若其相关总和为最大值，则 $\boldsymbol{X}_I$ 被选为基序列 $\mathrm{Base}_I$。

$$\mathrm{Base}_I = \left\{\boldsymbol{X}_I \middle| \mathrm{RelSum}_I = \max,\quad 0 < I \leq N\right\} \tag{6}$$

依据相关总和与相关次数最大值的方法，选取基序列，使 $\boldsymbol{S}$ 中基序列的相关序列的个数最多，利用基序列实现对其相关序列的拟合压缩，可提高压缩比和压缩精度。

#### 3.1.2 数据相关性分组

(1)相关序列分组：　若 $\boldsymbol{X}_I$ 为基序列，则与 $\boldsymbol{X}_I$ 相关的所有序列为其相关序列，$\boldsymbol{S}$ 中所有具有数据相关性的序列可以被划分成若干个相关集合 $\mathrm{RelSet}_I(I=1,2,\cdots,L)$。其存储格式为 $\mathrm{RelSet}_I = \{\mathrm{Base}_I, \boldsymbol{X}_J, \boldsymbol{X}_K, \cdots\}$，表示基序列 $\boldsymbol{X}_I$ 与其相关序列 $\boldsymbol{X}_J$, $\boldsymbol{X}_K$ 等的集合。

(2)独立序列分组：　$\boldsymbol{S}$ 中与其他任何序列均不相关的为独立序列。所有独立序列都被分到独立序列集 UniqSet 中。独立序列集的存储格式为 $\mathrm{UniqSet} = \{\boldsymbol{X}_R, \boldsymbol{X}_S, \boldsymbol{X}_T, \cdots\}$，表示此集合中所有序列之间是线性无关的。数据相关性分组算法的描述如表 1 所示。

**表 1　数据相关性分组算法**

| | |
|---|---|
| 算法输入： | 原始数据矩阵$\boldsymbol{S}=[\boldsymbol{X}_1,\boldsymbol{X}_2,\cdots,\boldsymbol{X}_N]$； |
| | 相关性阈值RelThreshold |
| 算法输出： | 相关序列集$\mathrm{RelSet}_I(I=1,2,\cdots,L)$； |
| | 独立序列集UniqSet |
| 算法步骤： | |
| 步骤 1 | for $i$=1: Column Num of $\boldsymbol{S}$ |
| | 计算$\mathrm{RelNum}_I$，建立候选基集合CanBase； |
| 步骤 2 | if(CanBase$\neq\phi$) |
| | (1)for $j$=1: Column Num of CanBase |
| | 计算$\mathrm{RelSum}_j$, $\mathrm{RelSum}_j$中值最大的被选为基序列$\mathrm{Base}_k$； |
| | (2)选取$\mathrm{Base}_k$的相关序列，得到相关集合$\mathrm{RelSet}_k$； |
| | (3)$\boldsymbol{S}$中除去$\mathrm{RelSet}_k$中的序列 |
| | goto 步骤1； |
| | else　得到独立序列集UniqSet； |
| 步骤 3 | 算法结束 |

### 3.2 基于序列相关性的两步数据压缩算法

簇内节点接收到来自簇首的分组信息后，执行分类压缩算法，通过计算拟合系数实现对相关序列的分类压缩，去除数据的属性相关性；对基序列和独立序列执行多项式分段拟合压缩，去除数据的时间相关性；将得到的压缩参数传至簇首，簇首根据压缩参数再次执行分类压缩算法。

**3.2.1 簇内节点的分类压缩算法**　首先，簇内节点针对相关序列执行拟合压缩算法。设某一相关序列集合为 $\mathrm{RelSet}_I = \{\mathrm{Base}_I, \boldsymbol{X}_J, \boldsymbol{X}_K\}$，令 $\boldsymbol{X}_J$ 和 $\boldsymbol{X}_K$ 的估计值分别为 $\overline{\boldsymbol{X}}_J = b_J + k_J\boldsymbol{X}_I$, $\overline{\boldsymbol{X}}_K = b_K + k_K\boldsymbol{X}_I$，只需分别计算出拟合系数 $(b_J, k_J)$ 与 $(b_K, k_K)$，即实现了对相关序列 $\boldsymbol{X}_J$ 和 $\boldsymbol{X}_K$ 的分类压缩。以 $\boldsymbol{X}_J$ 的求解为例，只需使目标函数 $Q = \sum_k (b_J + k_J x_{ik} - x_{jk})^2$ 的值最小，便可求出 $b_J$ 与 $k_J$，即令 $Q$ 对 $b_J$ 与 $k_J$ 的偏导数为 0 并求解方程组：

$$\left.\begin{aligned} &\sum_k \left(b_J + k_J x_{ik} - x_{jk}\right)\cdot\left(x_{ik}\right) = 0 \\ &\sum_k \left(b_J + k_J x_{ik} - x_{jk}\right) = 0 \end{aligned}\right\} \tag{7}$$

由此可得，$\boldsymbol{X}_J$ 中的每个监测值 $x_{jk}$ 的估计值 $\overline{x}_{jk}$，即 $\overline{x}_{jk} = b_J + k_J x_{ik}\,(k = 1,2,\cdots,m)$。

然后，簇内节点针对基序列和独立序列执行多项式拟合压缩算法。对时间序列的逼近求解，可采用多项式拟合方法去除数据的时间相关性。为了使算法能够较好地匹配无线传感器网络中数据的变化特点，本文提出了一种分段式的拟合压缩算法。

设某基序列为 $\boldsymbol{X}_I=(x_{i1},x_{i2},\cdots,x_{ik},\cdots,x_{im})^{\mathrm{T}}$，将 $\boldsymbol{X}_I$ 中的数据 $x_{ik}$ 利用最小二乘法加入到当前分段 $(x_{i1},x_{i2},\cdots,x_{ik})$，求出该段的拟合多项式 $p_n(x)=\sum_{i=0}^{n}a_ix^i$ ($n$ 为拟合次数)及均方误差值 RMSE。若 RMSE 值小于拟合误差限 ErrorThreshold，则将该段延长 $(x_{i1},x_{i2},\cdots,x_{ik},\cdots,x_{i(k+n)})$，重复上述过程求解，直至均方误差大于拟合误差限为止。该分段结束计算后，节点将该段时间起点 $T_{\mathrm{st}}=1$，终点 $T_{\mathrm{en}}=x_{i(k-1)}$ 和压缩参数(即为系数向量) $\mathrm{CoefVec}=(a_0,a_1,\cdots,a_n)$ 记录下来并存储，原来的压缩数据 $(x_{i1},x_{i2},\cdots,x_{i(k-1)})$ 将被丢弃，同时将 $x_{ik}$ 作为下一个分段的起点。 对独立序列的拟合方法与上述基序列拟合过程类似。算法实现描述如表 2 所示。

**3.2.2 簇首的分类压缩算法** 一方面，WSN 中相邻节点监测到的数据通常会表现出相同的变化趋势，这种相似性即为数据的空间冗余，簇首通过压缩方法能够将其去除；另一方面，原始采集数据也会存在一定误差，这些误差来自于节点自身特性或周围环境影响。簇首若能够对接收到的簇内数据进一步比较与整合，可以降低压缩误差，有效提高压缩效率和压缩精度。

**表 2 分段式拟合压缩算法**

| |
|---|
| 算法输入：原始数据$(x_{i1},1),(x_{i2},2),\cdots,(x_{im},m)$；<br>拟合多项式次数$n$；<br>拟合误差限ErrorThreshold |
| 算法输出：各分段的时间起点 $T_{\mathrm{st}}$与终点$T_{\mathrm{en}}$；<br>压缩参数CoefVec |
| 算法步骤： |
| 步骤 1 初始化$T_{\mathrm{st}}=1$； |
| 步骤 2 $T_{\mathrm{en}}=T_{\mathrm{st}}+n$；//初始段截止时间 |
| 步骤 3 if($T_{\mathrm{en}}\le m$) |
| (1)对$(T_{\mathrm{st}},T_{\mathrm{en}})$时间段内采集的数据拟合，得到拟合多项式 $p_n(x)$及压缩参数CoefVec； |
| (2)计算RMSE值； |
| (3)if( RMSE < ErrorThreshold ) |
| CoefVec = CoefVec ； |
| $T_{\mathrm{en}}=T_{\mathrm{en}}+1$； |
| //进行下一轮计算 |
| goto 步骤3； |
| else |
| $T_{\mathrm{en}}=T_{\mathrm{en}}-1$； |
| //当前段拟合结束 |
| 保存$T_{\mathrm{st}},T_{\mathrm{en}}$和压缩参数CoefVec； |
| (4)$T_{\mathrm{st}}=T_{\mathrm{en}}$ //下一分段开始 |
| goto 步骤2； |
| else 保存$(T_{\mathrm{st}},m)$时间段的数据； |
| 步骤 4 算法结束 |

假设一个簇内共有 $p$ 个节点。簇首在收到簇内节点的压缩参数之后，首先对其进行解压缩，得到节点 $k$ 中原始数据的逼近数据 $\overline{\boldsymbol{S}}(k)$：

$$\overline{\boldsymbol{S}}(k)=\left[\overline{\boldsymbol{X}}_1(k)\ \ \overline{\boldsymbol{X}}_2(k)\ \ \cdots\ \ \overline{\boldsymbol{X}}_N(k)\right]=\begin{bmatrix}\overline{x}_{11}(k) & \overline{x}_{21}(k) & \cdots & \overline{x}_{n1}(k)\\ \overline{x}_{12}(k) & \overline{x}_{22}(k) & \cdots & \overline{x}_{n2}(k)\\ \vdots & \vdots & \ddots & \vdots\\ \overline{x}_{1m}(k) & \overline{x}_{2m}(k) & \cdots & \overline{x}_{nm}(k)\end{bmatrix}_{M\times N} \tag{8}$$

为消除相邻节点采集数据之间的空间相关性，利用均值 $\mathrm{AVE}_{ij}$ 求出在某时刻上不同节点的采样均值，再对求得的均值进行分类压缩，压缩方法与前述簇内节点的压缩算法一致。此方法同时还能消除数据采集过程中存在的误差。

$$\mathrm{AVE}_{ij}=\frac{1}{p}\sum_{k=1}^{p}\overline{x}_{ij}(k) \tag{9}$$

## 4 算法分析与仿真

### 4.1 算法分析

算法中，簇内节点进行本地分类压缩，不执行分组算法，可节省部分计算能耗，分类压缩可降低簇内节点与簇首之间的通信成本，减少网络传输数据量。簇首先对监测数据分组，再对簇内数据压缩，由于簇首对本地数据分组后不再进行额外的压缩处理，因此也降低了簇首节点自身的计算能耗和通信能耗。因此文中提出的算法可以充分减少网络传输数据量，降低能量消耗，平衡了 WSN 中簇内节点与簇首的能量消耗，延长网络生命周期。

### 4.2 评价模型

压缩比、压缩误差及通信能耗是 WSN 压缩算法研究中的 3 个常见评价指标，但若算法较复杂，其计算能耗是不能被忽略的，甚至会出现超出所节省通信能耗的情况。数据压缩最主要的目的是尽量降低计算成本，减少通信能耗。因此从整体能耗的角度对算法进行评价具有更现实的意义。本文以一阶无线模型进行网络能耗分析[11]，提出了简化的基于能量判别的评估指标，即网络压缩能耗比(Network Compression Energy Ratio, NCER)，从整体能耗角度出发，并综合考虑基本的压缩比与压缩误差要求，来对算法进行综合性能评价。

**定义 4** 执行压缩算法后网络的总能耗 $E_{\mathrm{C}}$ 与压

缩前的总能耗 $E_{\mathrm{UC}}$ 的比值定义为网络压缩能耗比：

$$\mathrm{NCER} = E_{\mathrm{C}}/E_{\mathrm{UC}} \tag{10}$$

执行压缩算法前，网络的能耗主要包括：节点的休眠/唤醒能耗、数据采集能耗、无线通信能耗等。压缩后，除了上述 4 种能耗之外，还会存在计算能耗。其中，采集能耗、休眠/唤醒能耗在压缩前后基本不变，通常这 3 种能耗值非常低[15]，可忽略不计，由于执行了压缩算法，无线通信能耗必将有所降低。因此，对算法的性能评价只需考虑网络是否执行压缩算法时的通信能耗与计算能耗即可。由此可得：

$$\left.\begin{aligned} E_{\mathrm{UC}} &= E_{\mathrm{TX}}(L,d) + E_{\mathrm{RX}}(L) \\ E_{\mathrm{C}} &= E_{\mathrm{TX}}(L\times r(\delta),d) \\ &\quad + E_{\mathrm{RX}}(L\times r(\delta)) + E_{\mathrm{CPU}}(L,t) \end{aligned}\right\} \tag{11}$$

其中，$E_{\mathrm{TX}}$ 为发送能耗，$E_{\mathrm{RX}}$ 为接收能耗，$E_{\mathrm{CPU}}$ 为计算能耗，其值为 $E_{\mathrm{CPU}}=L\times t\times E'_{\mathrm{CPU}}$，$E'_{\mathrm{CPU}}$ 表示处理器运行功耗，$L$ 为原始数据长度，$d$ 为数据传输距离，$r$ 为压缩比，$\delta$ 为压缩误差，$t$ 为压缩单位比特数据的执行时间。

若网络中有 $N$ 个传感器节点，其中簇首为 $K$ 个，那么

$$\begin{aligned} \mathrm{NCER} = &\left[\sum_{i=1}^{N} E_{\mathrm{TX}}(L\times r_i(\delta),d_i)\right. \\ &\left.+\sum_{i=1}^{N-K} E_{\mathrm{RX}}(L\times r_i(\delta)) + \sum_{i=1}^{N} E_{\mathrm{CPU}}(L,t_i)\right] \\ &\Big/\sum_{i=1}^{N} E_{\mathrm{TX}}(L,d_i) + \sum_{i=1}^{N-K} E_{\mathrm{RX}}(L) \end{aligned} \tag{12}$$

式(12)表明，网络压缩能耗比 NCER 包含压缩比 $r$，压缩误差 $\delta$ 和计算能耗 $E_{\mathrm{CPU}}$。若 NCER<1，说明压缩后网络实现了节能，且 NCER 值越小，表明节能效果越好；若 $\mathrm{NCER}\geq 1$，说明压缩后网络未节省能耗。

### 4.3 仿真对比

本文通过 MATLAB7.0 平台对算法进行仿真分析，在 50 m×50 m 的区域内随机部署 100 个 SN。采用 LEACH 协议[16]对网络进行分簇并采用该协议中的能耗模型，簇首个数占网络中总节点个数的 4%。簇首与 Sink 节点之间的平均通信距离为 100 m。误差 $\delta$ 采用均方误差。仿真实验采用 $1024\times 4$ 个具有 4 种属性(温度、湿度、节点电压以及光强)的数据，来自于伯克利大学联合研究实验室利用 54 个 MicaZSN 采集的数据。拟合多项式次数 $n$ 取 1, 2, 3，相关性阈值 RelThreshold 分别为 0.7, 0.8, 0.9, 1.0 时，将本文所提 TSC-SC 算法与文献[14]提出的基于小波的自适应多模数据压缩(Adaptive Multiple-Modalities data Compression, AMMC)算法，文献[8]中提出的基于一元线性回归模型的空时数据压缩(One-Dimensional Linear Regression model based Spatial and Temporal, ODLRST)算法进行对比。分别从 3 个常用的评价指标：压缩比、压缩误差、通信能耗及本文提出的 NCER 评价模型对两种算法进行仿真对比。

**4.3.1 压缩比分析**　由图 1 可以看出：(1)TSC-SC 与 AMMC, ODLRST 算法的压缩比均随着相关性阈值的增大而增大，但 TSC-SC 算法的压缩比始终小于其他两种算法。分析其原因是相关性阈值 RelThreshold 越大，数据序列之间的相关性越弱，因此需要更多的参数来描述。当 RelThreshold=1 时，此时不考虑数据相关性，压缩比达到了最大值。ODLRST 算法采用对时间数据序列逐步判断添加方法，AMMC 算法采用了固定的 3 次小波分解压缩，而 TSC-SC 对数据分组算法进行了改进，采用分段式方法对多项式逼近拟合压缩，获得了较好的压缩效果。(2)RelThreshold 相同的情况下，拟合次数 $n$ 值越大，对数据序列分段个数越少，因此 TSC-SC 算法的压缩比越小。

**4.3.2 压缩误差分析**　由图 2 可以看出：(1)TSC-SC 算法的压缩误差始终低于 AMMC 和 ODLRST 算法，并且当 $n$ 取值不同时，拟合误差始终保持稳定。主要原因是 TSC-SC 算法进行拟合压缩时，是根据给定拟合误差限(ErrorThreshold)进行自适应的分段拟合，即 TSC-SC 算法是误差受限的分段拟合压缩，具有较好的收敛性。(2)当 RelThreshold 升高时，数据序列之间的相似度越强，压缩效果得到进一步提高，因此，均方误差 RMSE 值有所降低，拟合精度也就越高。

**4.3.3 通信能耗分析**　按照文中所提出的能耗模型进行仿真分析。由图 3 可以看出：(1)随着拟合次数 $n$ 的增加，TSC-SC 算法的压缩效果越好，通信能耗进一步降低，且始终低于另外两种算法。(2) RelThreshold 值越大，TSC-SC 的压缩比增加，所需的发送能耗增加，因此 TSC-SC 算法的通信能耗有所上升。

**4.3.4 压缩能耗比分析**　由图 4 可以看出：拟合次数 $n$=1 时，TSC-SC 算法的网络压缩能耗比 NCER 最低，节能效果也最好。$n$=2 时，虽然算法复杂度增加，但由于此时压缩比是降低的，因此也达到了较好的压缩效果。$n$ 为 3 时，TSC-SC 算法虽然在之前的比较中均表现出较好的性能，但由于算法较为复杂，因此其节能效果并不理想。由此可以得出结论：算法的计算能耗是不能被忽略的，用 NCER 指标能够更综合、直观地对算法进行性能评估。

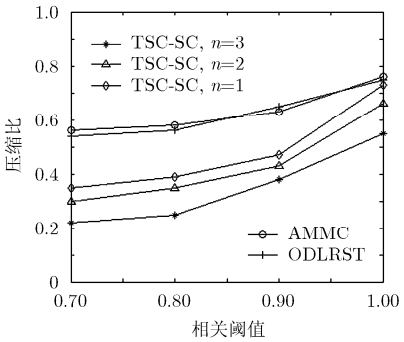


图 1 压缩比比较

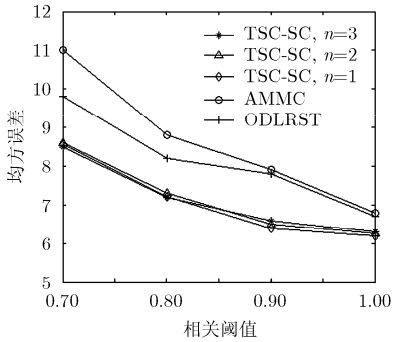


图 2 压缩误差比较

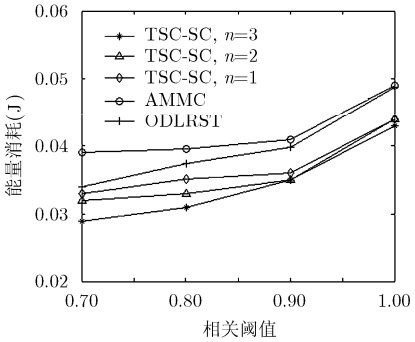


图 3 通信能耗比较

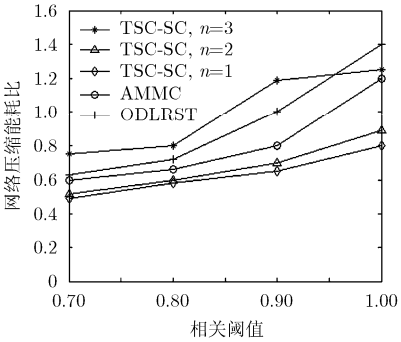


图 4 网络压缩能耗比(NCER)

在实际应用中，TSC-SC 算法的 1, 2 次拟合已经可以实现较好的数据压缩效果。因此，又进一步证明了 TSC-SC 算法的优越性能。

## 5 结束语

网络节能一直是无线传感器网络中要解决的关键问题之一。采用数据压缩技术可以有效减少网络中的传输数据量，降低网络能耗。文中基于 WSN 中数据的多维相关性，针对网络中的簇首和簇内节点，提出了两步压缩算法(TSC-SC)：簇首执行相关性数据分组算法，将数据分组，对接收到的来自簇内节点的压缩参数解压缩，利用均值函数求得采样均值后，再次进行分类压缩，充分消除了数据的空间相关性。簇内节点对多属性数据分类压缩，去除属性相关性；利用多项式拟合方法去除时间相关性，为了更好地匹配 WSN 中数据多变的特点，提出一种分段式的拟合压缩算法，达到了较好的压缩效果。为了综合地评价压缩算法的性能，提出了一种基于能量判别的评估指标，综合考虑了压缩比、压缩误差、通信能耗和计算能耗，对压缩算法的性能直观地进行评价，具有实际意义。

## 参 考 文 献

[1] 唐亮, 周正, 石磊, 等. 基于能量均衡的无线传感器网络压缩感知算法[J]. 电子与信息学报, 2011, 33(8): 1919–1923. doi: 10.3724/SP.J.1146.2010.01388.
TANG L, ZHOU Z, SHI L, *et al*. Energy balance based WSN compressive sensing algorithm[J]. *Journal of Electronics and Information Technology*, 2011, 33(8): 1919–1923. doi: 10.3724/SP.J.1146.2010.01388.

[2] KIM H W, SEO H S, HONG S H, *et al*. Modeling of energy-efficient applicable routing algorithm in WSN[J]. *International Journal of Digital Content Technology & Its Applications*, 2010, 4(5): 13–22.

[3] ERRATT N and LIANG Y. Compressed data-stream protocol: an energy-efficient compressed data-stream protocol for wireless sensor networks[J]. *IET Communications*, 2011, 5(18): 2673–2683. doi: 10.1049/iet-com.2011.0118.

[4] 钱志鸿, 王义君. 面向物联网的无线传感器网络综述[J]. 电子与信息学报, 2013, 35(1): 215–227. doi: 10.3724/SP.J.1146.2012.00876.
QIAN Zhihong and WANG Yijun. Internet of things-oriented wireless sensor networks review[J]. *Journal of Electronics & Information Technology*, 2013, 35(1): 215–227. doi: 10.3724/SP.J.1146.2012.00876.

[5] 徐晓滨, 文成林, 刘荣利. 基于随机集理论的多源信息统一表示与建模方法[J]. 电子学报, 2008, 36(6): 1174–1181.
XU Xiaobin, WEN Chenglin, and LIU Rongli. The unified method of describing and modeling multisource information based on random set theory[J]. *Acta Electronica Sinica*, 2008,

36(6): 1174–1181.

[6] 李阳阳, 王洪波, 张鹏, 等. 基于多属性信息的数据中心间数据传输调度方法[J]. 通信学报, 2012, 33(Z1): 121–131. doi: 10.3969/j.issn.1000-436x.2012.z1.016.
LI Yangyang, WANG Hongbo, ZHANG Peng, *et al*. Multi-attribute aware scheduling for inter-datacenter bulk transfers[J]. *Journal on Communications*, 2012, 33(Z1): 121–131. doi: 10.3969/j.issn.1000-436x.2012.z1.016.

[7] CHOU J, PETROVIC D, and RAMCHANDRAN K. A distributed and adaptive signal processing approach to exploiting correlation in sensor networks[J]. *Ad Hoc Networks*, 2004, 2(4): 387–403. doi: 10.1016/j.adhoc.2003.09.001.

[8] 王雷春, 马传香. 传感器网络中一种基于一元线性回归模型的空时数据压缩算法[J]. 电子与信息学报, 2010, 32(3): 755–758. doi: 10.3724/SP.J.1146.2009.00704.
WANG Leichun and MA Chuanxiang. A one-dimensional linear regression model based spatial and temporal data compression algorithm for wireless senor networks[J]. *Journal of Electronics & Information Technology*, 2010, 32(3): 755–758. doi: 10.3724/SP.J.1146.2009.00704.

[9] 蒋鹏, 吴建峰, 吴斌, 等. 基于自适应最优消零的无线传感器网络数据压缩算法研究[J]. 通信学报, 2013, 34(2): 1–7. doi: 10.3969/j.issn.1000-436x.2013.02.001.
JIANG Peng, WU Jianfeng, WU Bin, *et al*. Data compression method for wireless sensor networks based on adaptive optimal zero suppression[J]. *Journal on Communications*, 2013, 34(2): 1–7. doi: 10.3969/j.issn.1000-436x.2013.02.001.

[10] 杜卓明, 耿国华, 贺毅岳. 一种基于压缩感知的二维几何信号压缩方法[J]. 自动化学报, 2012, 38(11): 1841–1846. doi: 10.3724/SP.J.1004.2012.01841.
DU Zhuoming, GENG Guohua, and HE Yiyue. A 2-D geometric signal compression method based on compressed sensing[J]. *Acta Automatic*, *Sinaca*, 2012, 38(11): 1841–1846. doi: 10.3724/SP.J.1004.2012.01841.

[11] 陈正宇, 杨庚, 陈蕾, 等. 基于压缩感知的 WSNs 长生命周期数据收集方法[J]. 电子与信息学报, 2014, 36(10): 2343–2349. doi: 10.3724/SP.J.1146.2013.01787.
CHEN Zhengyu, YANG Geng, CHEN Lei, *et al*. Data gathering for long network lifetime in WSNs based on compressed sensing[J]. *Journal of Electronics & Information Technology*, 2014, 36(10): 2343–2349. doi: 10.3724/SP.J.1146.2013.01787.

[12] 黄海平, 陈九天, 王汝传, 等. 无线传感器网络中基于数据融合树的压缩感知算法[J]. 电子与信息学报, 2014, 36(10): 2364–2369. doi: 10.3724/SP.J.1146.2013.01621.
HUANG Haiping, CHEN Jiutian, WANG Ruchuan, *et al*. Compressed sensing algorithm based on data fusion tree in wireless sensor networks[J]. *Journal of Electronics & Information Technology*, 2014, 36(10): 2364–2369. doi: 10.3724/SP.J.1146.2013.01621.

[13] 周四望, 李兰. 传感器网络基于 DTW 的多小波压缩算法[J]. 通信学报, 2014, 35(8): 86–94. doi: 10.3969/j.issn.1000-436x.2014.08.012.
ZHOU Siwang and LI Lan. DTW-based multi-wavelet data compression algorithm for wireless sensor networks[J]. *Journal on Communications*, 2014, 35(8): 86–94. doi: 10.3969/j.issn.1000-436x.2014.08.012.

[14] 朱铁军, 林亚平, 周四望, 等. 无线传感器网络中基于小波的自适应多模数据压缩算法[J]. 通信学报, 2009, 30(3): 48–53.
ZHU Tiejun, LIN Yaping, ZHOU Siwang, *et al*. Adaptive multiple-modalities data compression algorithm using wavelet for wireless sensor networks[J]. *Journal on Communications*, 2009, 30(3): 48–53.

[15] POLASTRE J, SZEWCZY K R, and CULLER D. Telos: enabling ultra-low power wireless research[C]. Proceedings of the 4th International Symposium on Information Processing in Sensor Networks, California, 2005: 364–369. doi: 10.1109/IPSN.2005.1440950.

[16] HEINZELMAN W, CHANDRAKASAN A, and BALAKRISHNAN H. Energy-efficient communication protocol for wireless micro sensor networks[C]. Proceedings of the 33rd Annual Hawaii International Conference on System Sciences, Hawaii, 2000: 3005–3014.

翟　双：　女，1978 年生，讲师，研究方向为无线通信网络关键技术研究.

钱志鸿：　男，1957 年生，教授，研究方向为无线通信网络与通信系统研究.

刘晓慧：　女，1989 年生，硕士生，研究方向为无线通信网络关键技术研究.