

# 传感器网络基于 DTW 的多小波压缩算法

周四望, 李兰

(湖南大学 信息科学与工程学院, 湖南 长沙 410082)

**摘要:** 提出传感器网络环境下基于 DTW 的多小波数据压缩算法。首先研究汇聚节点中异步数据点一对一对的对应关系, 设计迭代算法求出具有最大相关性的 DTW 弯曲路径。接着提出最佳匹配点选择算法, 通过 DTW 弯曲路径中一对一数据点一对对来预测异步数据向量间的函数关系, 获取最佳匹配点, 得到具有最大相关性的传感数据矩阵。然后设计多小波变换, 利用传感数据矩阵的相关性来压缩数据, 同时解决数据矩阵的行列不对称问题。实验结果表明, 所提出的算法在能量聚集比、重构精度和运行时间等压缩性能指标上优于经典的分布式小波压缩算法。

**关键词:** 传感器网络; DTW; 小波; 数据压缩

中图分类号: TP393

文献标识码: A

文章编号: 1000-436X(2014)08-0086-09

## DTW-based multi-wavelet data compression algorithm for wireless sensor networks

ZHOU Si-wang, LI Lan

(College of Computer Science and Electronic Engineering, Hunan University, Changsha 410082, China)

**Abstract:** A data compression algorithm for wireless sensor networks based on DTW and multi-wavelet transform is proposed. Firstly, the correlation and warping path of asynchronous data point pairs is introduced, and then an iterative algorithm for solving DTW warping path with maximal similarity is designed. Secondly, a best match point algorithm is presented, which employs prediction to estimate the relationship of synchronous data vectors and then find out the best match points. A sensory data matrix with maximal correlation is thus obtained. Afterwards, a multi-wavelet transform is introduced, which is capable of utilizing the spatial correlation of sensory data matrix and solving the row-column asymmetry problem. Experiment results show that proposed method has higher energy concentration ratio, better reconstruction accuracy and lower running time compared to the traditional distributed wavelet compression algorithm.

**Key words:** wireless sensor network; DTW; wavelet; data compression

### 1 引言

无线传感器网络是无线网络重点研究的内容, 其应用背景已经从单纯的军事领域逐渐转移到环境监测、病人监护、空间探索等诸多领域。在传感器网络的生命周期内, 数量众多的传感器产生了海量的感知数据流。然而, 传感器网络的计算、存储、能量等资源受限, 难以承载大数据量的传感数据。数据压缩技术能够依据用户对数据精度的要求大

幅度减少网内数据量, 也能在相应的精度范围内重构原始传感数据。传感器网络数据压缩算法因此正吸引着越来越多研究人员的关注<sup>[1,2]</sup>。

小波压缩是目前传感器网络数据压缩研究的重点。在传感器网络中, 单个节点产生的传感数据在时间上是相关的, 地理位置相邻的节点收集到的传感数据在空间上往往也是相关的。有着“数学显微镜”之称的小波变换具有很好的去相关特性, 特别适合压缩时空相关的传感数据。美国加州大学

收稿日期: 2013-03-28; 修回日期: 2013-08-15

基金项目: 国家自然科学基金资助项目(60973127); 新世纪优秀人才支持计划基金资助项目(NCET-11-0136); 湖南省自然科学基金资助项目(14JJ2051)

**Foundation Items:** The National Natural Science Foundation of China(60973127); Program for New Century Excellent Talents in University (NCET-11-0136); The Natural Science Foundation of Hunan Province (14JJ2051)

洛杉矶分校(UCLA)研究了一种基于小波路由的 DIMENSIONS 算法,位于路由最底层的传感器节点首先对数据进行小波变换,降低时间冗余;簇头(汇聚节点)收集簇内节点传送的小波系数,进行联合小波压缩,降低空间冗余,大大减少了参与网内传输的数据量<sup>[3]</sup>。可以说, DIMENSIONS 建立了传感器网络小波数据压缩研究的基本框架。在传感器网络中,节点随机部署,因此节点的分布是不规则的。在不规则的传感器网络拓扑上直接运用小波变换难以取得好的数据压缩性能。据此,文献[4]提出了一种不规则提升小波变换以适应传感器网络不规则特性,取得了很好的压缩效果。从图论的观点来看,传感器网络可以看成是一个图,节点是图的顶点,而通信与处理开销是图的边。Ortega 等人因此展开了传感器网络图小波变换的研究,将小波变换和数据路由结合起来,形成图小波变换,提高了小波压缩的效率<sup>[5]</sup>。文献[6]研究传感器网络有损小波压缩问题,提出了基于 Haar 小波变换的动态小波摘要数据管理方法,不传输原始数据,从而降低了参与网络传输的数据量。文献[7]以图像数据为例研究传感器网络分布式小波变换,在提高压缩比的同时,能够降低节点的能量和存储开销。本文将小波函数融入传感器网络,提出了任意支撑长度的小波变换、时空小波变换、存储有效的小波变换和误差有界的小波变换<sup>[8]</sup>,取得了较高的数据压缩比。

从上述研究现状来看,现在成果主要集中在小波算法上,意图研究适合特定传感器网络环境的小波变换,从而提高小波压缩算法的效率。

然而,小波数据压缩是一种自适应的压缩方法,其压缩效率不仅与小波变换本身有关,还依赖于待压缩的数据。在传感器网络环境下,传感数据呈现异步性特征<sup>[9]</sup>。如何从传感数据固有的异步性角度来研究传感器网络数据压缩算法,还鲜有文献涉及。

本文基于动态时间弯曲(DTW, dynamic time warping)技术和多小波理论来研究传感器网络中的数据压缩问题。针对传感数据的异步性特征,在数据压缩之前先进行 DTW 预处理,获取具有最大相关性的数据矩阵;然后提出一种多小波变换,利用多小波的矢量性质来处理此矩阵行列不对称问题,形成基于 DTW 的多小波数据压缩算法。

## 2 预备知识

### 2.1 DTW 技术

DTW 是一种将距离测度与时间规整相结合的非线性测量技术<sup>[10]</sup>,其基本思想是在时间轴上将未知量不均匀地扭曲或者弯折,以使其特征与参考模式匹配。假设有 2 个时间序列  $Q$  和  $C$ ,其数据长度分别是  $N$  和  $M$

$$Q = q_1, q_2, \dots, q_n, \dots, q_N$$

$$C = c_1, c_2, \dots, c_m, \dots, c_M$$

将上述 2 个时间序列  $Q$  和  $C$  分别置于二维坐标的两轴,  $q_n$  和  $c_m$  在坐标平面上形成交点,设为  $w_k$ ,则  $Q$  和  $C$  对应的交点集合为 DTW 弯曲路径

$$w = w_1, w_2, \dots, w_k, \dots, w_K$$

DWT 弯曲路径满足以下 4 个条件。

- 1) 有界条件:  $\max(M, N) \leq K \leq M + N - 1$ ;
- 2) 边界条件:  $w_1 = (q_1, c_1)$ ,  $w_K = (q_N, c_M)$ ;
- 3) 连续条件: 给定  $w_k = (q_a, c_b)$  和  $w_{k-1} = (q_{a'}, c_{b'})$ , 必须有  $a - a' \leq 1$  且  $b - b' \leq 1$ ;
- 4) 单调条件: 给定  $w_k = (q_a, c_b)$  和  $w_{k-1} = (q_{a'}, c_{b'})$ , 必须有  $a - a' \geq 0$  且  $b - b' \geq 0$ 。

DTW 技术目前主要应用于模式识别领域,本文将借鉴其思想研究传感数据向量间的内在关系,寻找最具相关性的传感数据矩阵。

### 2.2 多小波

多小波和矢量滤波器对应,是小波理论研究的重要进展。多小波将经典多分辨分析空间扩展为  $r$  重多分辨分析,弥补了单小波必须在正交性和对称性之间进行折衷的不足。

设函数  $\phi = (\phi_1, \phi_2, \dots, \phi_r)^T \in L^2(R)$ ,  $r \geq 1$  为整数,对  $j \in Z$ , 定义  $V_j = \text{span}\{2^{j/2} \phi_i(2^j - k), 1 \leq i \leq r, k \in Z\}$ , 若空间序列  $V_j$  满足下列条件。

- 1)  $V_j \subset V_{j+1}, \forall j \in Z$ ;
- 2)  $\bigcup_{j \in Z} V_j = L^2(R)$ ,  $\bigcap_{j \in Z} V_j = \{0\}$ ;
- 3)  $f(\cdot) \in V_j \Leftrightarrow f(2\cdot) \in V_{j+1}, \forall j \in Z$ ;
- 4)  $f(x) \in V_j \Leftrightarrow f(x + \frac{1}{2^j}) \in V_{j+1}, \forall j \in Z$ ;
- 5)  $\{\phi_i(\cdot - k) : 1 \leq i \leq r, k \in Z\}$  是  $V_0$  空间 Riesz 的基。

则称  $\phi$  生成  $L^2(R)$  的  $r$  重多分辨分析  $MRA(V_j)$ 。

本文将利用多小波的  $r$  重矢量性质来处理具有行列不对称性质的传感数据矩阵。

### 3 基于 DTW 的多小波压缩算法

本节将借鉴 DTW 的思想研究异步传感数据的相关性, 通过设计最小距离弯曲路径计算算法和最佳匹配点选择算法, 获取最具相关性的传感数据矩阵。在此基础上研究多小波变换, 形成基于 DTW 的多小波数据压缩(DMC, DTW-based multi-wavelet compression)算法。

#### 3.1 传感数据的异步性

传感器网络中的数据具有异步特性, 这来源于传感器节点异步节能唤醒机制和多跳通信延迟。设环境监测传感器网络中有 2 个传感器节点  $S_i$  和  $S_j$  向汇聚节点  $S$  传送时间序列数据, 如图 1 所示。

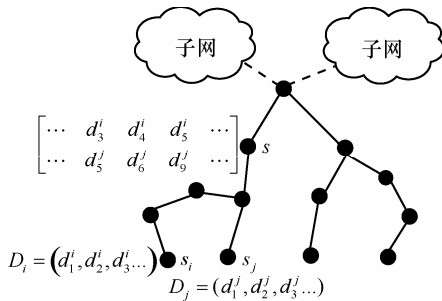


图 1 传感器网络数据传送

$S$  接收  $S_i$  和  $S_j$  传送的数据向量  $D_i$  和  $D_j$ , 组成传感数据矩阵  $D$

$$D = \begin{bmatrix} D_i \\ D_j \end{bmatrix} = \begin{bmatrix} \dots & d_{k+3}^i & d_{k+4}^i & d_{k+5}^i & \dots \\ \dots & d_{k+5}^j & d_{k+6}^j & d_{k+9}^j & \dots \end{bmatrix}$$

由于  $S_i$ 、 $S_j$  到  $S$  的通信延迟以及网络中可能出现的分组丢失情况( $d_{k+7}^j$  和  $d_{k+8}^j$  丢失),  $D$  中的数据在时序上不再具有对应关系, 具有异步性。

图 2 直观地给出了  $D_i$  和  $D_j$  的异步性关系。 $d_1^i$  和  $d_1^j$  同步, 但随后  $(d_2^i, d_2^j)$ 、 $(d_3^i, d_3^j)$ 、 $(d_4^i, d_4^j)$  等数据点一对的相似性逐渐减弱, 呈现异步性特征。小波压缩的效率取决于小波变换的去相关性程度。

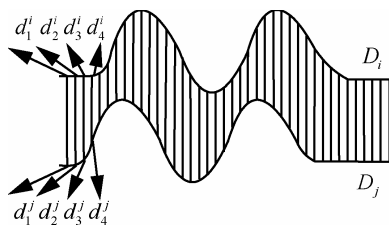


图 2 传感数据的异步性

因此, 直接对  $D$  进行小波变换很难取得好的压缩性能。事实上,  $D_i$  只是略微滞后于  $D_j$ , 从总体上看  $D_i$  和  $D_j$  是很相似的。如何分析异步数据的内在相似性, 提出相应的小波压缩算法是本文重点研究的内容。

#### 3.2 最小距离弯曲路径算法

设有一组传感器节点向汇聚节点传输传感数据, 其中第  $i$  和第  $j$  个节点传送的传感数据向量分别为  $D_i$  和  $D_j$ 。设  $D_i$  和  $D_j$  的数据长度分别为  $M$  和  $N$ 。

**定义 1** (距离矩阵): 在传感器网络中, 传感数据向量  $D_i$  和  $D_j$  的距离矩阵  $D\_matrix$  是一个  $N$  行  $M$  列的矩阵, 矩阵元素  $d(d_m^i, d_n^j)$  为数据点  $d_m^i$  和  $d_n^j$  之间的欧几里德距离, 即:  $d(d_m^i, d_n^j) = |d_m^i - d_n^j|$ 。

依据定义 1, 距离矩阵  $D\_matrix$  具体表示为

$$D\_matrix = \begin{bmatrix} d(d_1^i, d_N^j) & d(d_2^i, d_N^j) & \dots & d(d_M^i, d_N^j) \\ d(d_1^i, d_{N-1}^j) & d(d_2^i, d_{N-1}^j) & \dots & d(d_M^i, d_{N-1}^j) \\ \vdots & \vdots & \ddots & \vdots \\ d(d_1^i, d_1^j) & d(d_2^i, d_1^j) & \dots & d(d_M^i, d_1^j) \end{bmatrix}$$

在  $D\_matrix$  中, 元素  $d(d_m^i, d_n^j)$  可以看作是数据点  $d_m^i$  和  $d_n^j$  之间相关性的量化表示。 $d_m^i$  和  $d_n^j$  越相关, 则  $d(d_m^i, d_n^j)$  的值就越趋近于 0; 相应地, 若  $d(d_m^i, d_n^j)$  的值增大, 则表明  $d_m^i$  和  $d_n^j$  越相异。

将传感数据向量  $D_i$  和  $D_j$  分别置于二维坐标的两轴, 如图 3 所示, 进而可以给出  $D_i$  和  $D_j$  弯曲路径的定义。

**定义 2** (弯曲路径): 在传感数据向量  $D_i$  和  $D_j$  的距离矩阵  $D\_matrix$  中, 满足 DTW 有界性、边界条件、连续性和单调性的一组连续的矩阵元素集合称为一条弯曲路径。

图 3 画出了一条弯曲路径  $w = w_1, w_2, w_3, \dots, w_K$ 。路径  $w$  中的每一点对应距离矩阵  $D\_matrix$  的一个元素。经典 DTW 理论从数学意义上给出了弯曲路径的有界性、边界条件、连续性和单调性等限制条件的定义。对于本文研究的传感数据弯曲路径问题, 有界性要求弯曲路径中元素总数  $K$  必须满足  $\max(M, N) \leq K \leq M + N - 1$ ; 边界条件确保距离矩阵斜对角线上的两端元素就是弯曲路径的起止元素, 即  $w_1 = d(d_1^i, d_1^j)$  且  $w_K = d(d_M^i, d_N^j)$ ; 连续性约束路径中的每一个弯折步必须彼此相邻, 即路径是不间断的; 而单调性强制保证弯曲路径在时间轴上是单调的, 即不允许  $w$  中的点回退。

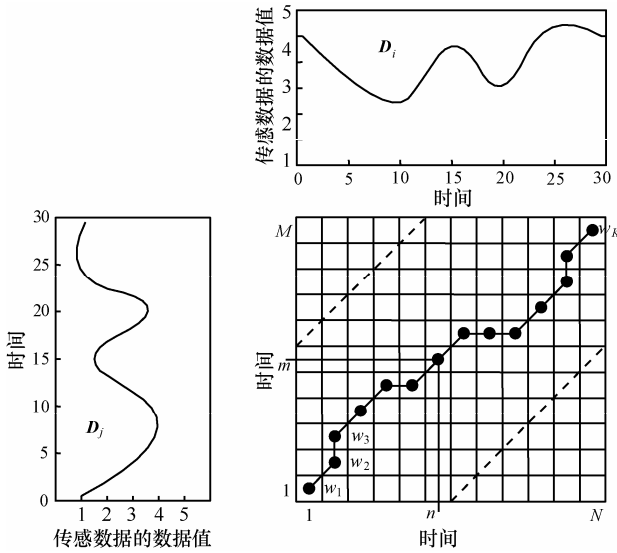


图 3 DTW 弯曲路径

图 4 是依据 DTW 弯曲路径获取的数据点一点对关系。在图 4 中，利用距离矩阵初步处理了传感数据的异步性。显然，与图 2 相比，图 4 更能体现数据向量  $D_i$  和  $D_j$  的相关性。

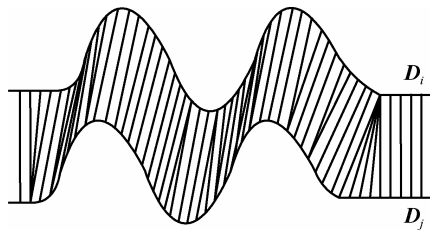


图 4 与弯曲路径对应的数据点一点对

从弯曲路径的定义可以看出，理论上满足 DTW 时间弯曲限制条件的弯曲路径不止一条，存在多解。弯曲路径越长，数据量就越大，不利于数据压缩。而对于同样长度的弯曲路径来说，数据点相关性越高的弯曲路径，小波压缩的效率越高，因而需要寻找具有最大相关性的弯曲路径。

**性质 1** 在满足 DTW 限制条件的弯曲路径中，具有最小距离弯曲路径的数据点一点对存在最大相关性。

在定义 1 中，数据相关性被量化为数据点一点对之间的欧几里德距离，距离越小则相关性越大。而由定义 2 可知，满足 DTW 限制条件的弯曲路径有多条。显然，数据点一点对距离最小的那条弯曲路径具有最大的相关性。

设  $DWT(D_i, D_j)$  表示具有最大相关性的弯曲路径距离。则依据性质 1， $DWT(D_i, D_j)$  可表示为

$$DWT(D_i, D_j) = \min_l \left\{ \sum_{k=1}^K \frac{w_k^l}{K} \right\}$$

其中， $K$  起规范化作用，便于不同长度弯曲路径之间的比较。 $w_k^l$  指第  $l$  条弯曲路径上的第  $k$  个数据点。理论上，可以利用穷举搜索法找出所有满足条件的弯曲路径，并算出每条弯曲路径的动态弯曲距离，求出最小距离弯曲路径。但穷举搜索的计算量大，不适合资源受限的传感器网络。由优化理论可知，对于数据向量  $D_i$  和  $D_j$ ，设有距离矩阵元素  $d(d_m^i, d_n^j)$  在最佳弯曲路径上，那么从点  $(d_1^i, d_1^j)$  到  $(d_m^i, d_n^j)$  的子路径也是局部最优解。这样，局部最优解可通过迭代搜索获得。当  $D_i$  和  $D_j$  的点都计算完毕，就得到了具有最小距离的弯曲路径。最小距离弯曲路径迭代计算的具体算法如下。

初始化步： $m=1; n=1; \gamma = d(d_1^i, d_1^j)$

迭代步(终止条件： $m=M$  且  $n=N$ ):

- ①  $\gamma = \gamma + \min \{ d(d_x^i, d_y^j) \}_{(x=m+1, y=n); (x=m+1, y=n+1); (x=m, y=n)}$ ;
- ②  $m = x; n = y$ 。

最小距离弯曲路径迭代算法由初始化步和迭代步组成。在迭代步，算法只需比较 3 个邻近数据点一点对的距离，而无需获取距离矩阵  $D\_matrix$  的全局信息。因此，在 DTW 弯曲路径计算过程中，不需要构造完整的距离矩阵，算法随着数据的读入即时运行，节省了传感器网络有限的计算和存储资源。

### 3.3 最佳匹配点选择算法

在 3.2 节中，提出一种迭代算法，求出了具有最小距离的弯曲路径。对于传感数据向量  $D_i$  和  $D_j$ ，其弯曲路径存在如下几种数据“点一点”对。

- 1) 一对一：指  $D_i$  中的某一数据点和  $D_j$  中的某一数据点一一对应。
- 2) 一对多：指  $D_i$  中的某一数据点与  $D_j$  中的多个数据点对应。由 DTW 约束条件可知  $D_j$  中的这几个数据点在时间上是连续的。
- 3) 多对一：指  $D_i$  中的多个数据点与  $D_j$  中的一个点对应。同样，依据 DTW 约束条件可知  $D_i$  中的这几个数据点在时间上也是连续的。

**定义 3 (匹配点)**：对于传感数据向量  $D_i$  和  $D_j$

$$D_i = \{d_1^i, d_2^i, \dots, d_m^i, \dots, d_M^i\}$$

$$D_j = \{d_1^j, d_2^j, \dots, d_n^j, \dots, d_N^j\}$$

若  $D_i$  中的某个数据点  $d_m^i$  在  $D_j$  上仅有一个数

据点  $d_n^j$  与之对应, 则称  $d_n^j$  为  $d_m^i$  的匹配点。

由定义 3 可知, 存在一对一关系的数据点一点对互为匹配点。然而, 若  $D_i$  中的某个点与  $D_j$  中的点成一对多的关系, 则必须在  $D_j$  中选择最合适的一个数据点与之对应, 这便是最佳匹配点选择问题。

本节运用预测方法确定最佳匹配点。对于  $D_i$  和  $D_j$ , 由 3.2 节提出的迭代算法可求出二者的最小距离弯曲路径  $w$ ,  $w$  由对应的距离矩阵元素组成。首先将  $D_i$  中的元素看成是  $D_j$  中元素的函数, 即  $d_m^i = f(d_n^j)$ 。然后根据弯曲路径  $w$  上已有的一对一数据点一点对来拟合函数  $f(\cdot)$ , 即用已知来预测未知。拟合函数  $f(\cdot)$  表示  $D_i$  和  $D_j$  的变换关系。对存在一对多关系的数据点  $d_m^i$ , 通过拟合函数  $f(\cdot)$  求出理想的一对一匹配点  $d_x^j$ , 即  $d_x^j = f^{-1}(d_m^i)$ 。将  $d_x^j$  和弯曲路径中同  $d_m^i$  呈一对多关系的各数据点进行比较, 找出与之最为接近的数据点即为  $d_m^i$  的最佳匹配点。本文采用最小二乘法来拟合函数  $f(\cdot)$ , 最佳匹配点选择算法如下。

**Step1** 根据时间弯曲路径  $w$  初始化一对一点点对集合  $A$  和需要匹配的对象集合  $B$

$$A = \{(d_m^i, d_n^j) \mid d_m^i \in D_i, d_n^j \in D_j\}$$

$$B = \{(d_m^i, X) \mid d_m^i \in D_i, X \subseteq D_j\}$$

**Step2**  $d_m^i = f(d_n^j) = ad_n^j + b$ , 依据集合  $A$  并运用最小二乘法拟合计算参数  $a$  和  $b$  的值。

**Step3** 对集合  $B$  中的每一个元素  $d_m^i$ :

① 求  $d_x^j = f^{-1}(d_m^i) = (d_m^i - b) / a$ ,  $d_x^j$  即为与  $d_m^i$  对应的理想匹配点预测值;

② 将  $d_x^j$  与集合  $B$  中的  $X$  进行比较, 找出与其差值最小的点, 即为  $d_m^i$  的最佳匹配点。

可以看出, 上述最佳匹配点选择算法的实质是根据已有的一对一点一点对来预测 2 个数据向量的函数关系, 再由此函数从一对多点一点对中找出一个最佳匹配点, 形成新的一对一和多对一的对应关系。显然,  $D_i$  和  $D_j$  的多对一关系相当于  $D_j$  和  $D_i$  中的一对多关系。不难证明, 由最佳匹配点选择算法确定的新数据点一点对符合 DTW 有界性、边界条件、连续性和单调性等约束条件。这样, 通过运行最佳匹配点选择算法, 最终可以得到数据向量的一对一点一点对, 形成最具相关性的待压缩传感数据矩阵。

### 3.4 多小波变换

从小波理论的角度看, 目前应用于传感器网络的小波变换是一种单小波变换, 很难获得好的传感数据压缩性能。在传感器网络中, 汇聚节点常常将数据组织成二维数据矩阵。该矩阵每一行的数据间存在时间相关性, 行与行之间有空间相关性。对二维数据来说, 单小波变换分离为行小波变换和列小波变换, 即二维单小波变换。行-列小波变换交替执行, 形成多级小波变换。然而, 传感数据矩阵不同于图像等二维数据, 行数很少而列数很多, 即行和列明显的不对称, 行-列小波变换难以交替执行, 空间相关性得不到处理, 因而很难取得好的压缩效果。

在 3.2 节和 3.3 节, 以两数据向量  $D_i$  和  $D_j$  为例研究 DTW 弯曲路径问题, 找到了最具相关性的数据点对应关系, 克服了传感器数据集的异步性。上述结果很容易推广到由  $r$  个传感器节点产生的矢量数据集  $(D_1, D_2, \dots, D_r)^T$ 。本节研究针对矢量传感数据集的多小波变换方法, 利用  $r$  个数据向量之间的相关性来实现数据压缩。

对  $r$  个数据向量  $D_1, D_2, \dots, D_r$  执行弯曲路径距离迭代算法和最佳匹配点选择算法, 可以得到  $r$  个数据向量的一对一点一点对, 如图 5 所示。

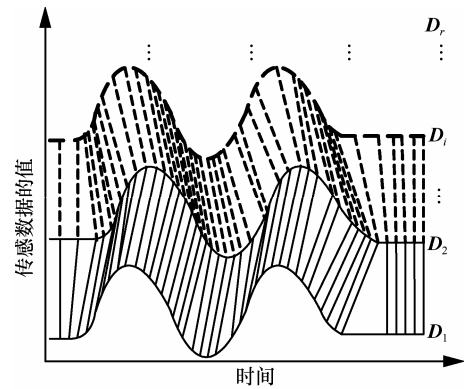


图 5  $D_1, D_2, \dots, D_r$  数据点一点对

汇聚节点将此  $r$  个数据点一点对组织成一个二维传感数据矩阵  $D$

$$D = \begin{pmatrix} D_1 \\ D_2 \\ \vdots \\ D_r \end{pmatrix} = \begin{pmatrix} d_1^1 & d_2^1 & \dots & d_r^1 \\ d_1^2 & d_2^2 & \dots & d_r^2 \\ \vdots & \vdots & \ddots & \vdots \\ d_1^r & d_2^r & \dots & d_r^r \end{pmatrix}$$

对二维数据来说, 经典的单小波变换分离为行

变换和列变换，即二维小波变换。行、列小波变换交替执行  $L$  次，形成  $L$  级小波变换，数据集被变换为低频部分以及水平高频、垂直高频和对角高频部分。图 6(a)是三级二维小波变换示例，其中  $A_3$  代表低频小波系数， $H_1$ 、 $H_2$ 、 $H_3$  分别表示 3 水平高频系数， $V_1$ 、 $V_2$ 、 $V_3$  和  $D_1$ 、 $D_2$ 、 $D_3$  表示相应的垂直和对角高频系数。图像是典型的二维数据，一般来说，图像数据的行数和列数对称，适合交替执行行、列小波变换。事实上，基于二维小波变换的图像压缩是小波理论在信息处理领域中最为成功的应用。

然而，传感数据矩阵  $D$  同于图像数据，并不适合采用图 6(a)所示的经典二维小波变换。

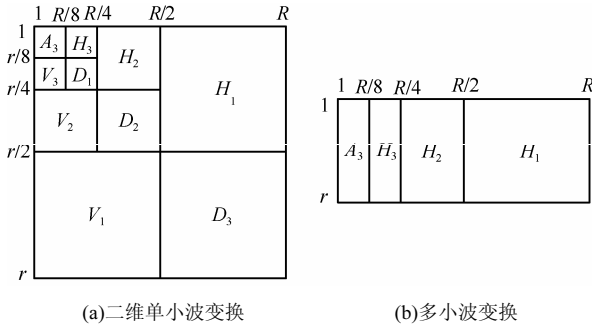


图 6 三级小波变换

**性质 2**  $D$  的行数少，但列数多，即  $r \ll R$ 。

性质 2 表明，数据集  $D$  具有行列不对称的性质。这种性质是由传感器网络特性决定的。经典的 LEACH 路由协议<sup>[11]</sup>指出，在传感器网络中最理想的簇头或汇聚节点的数量不应大于传感器节点总数的 5%。也就是说，簇的规模以不超过 20 个节点为最佳，即  $r \leq 20$ 。而  $R$  的物理意义是节点采集的数据量。若传感器节点采样频率为 1 kHz，以连续采集 5 s 时间为例，节点收集到数据将超过 5 000，即  $R > 5\ 000$ ，因此有  $r \ll R$ 。

传感数据集  $D$  是一个二维矩阵，但行数  $r$  远小于列数  $R$ ，因而需要研究新的小波变换方法。

针对  $r$  行二维矢量数据，设多尺度函数  $\phi = (\phi_1, \phi_2, \dots, \phi_r)^T$  生成平方可积空间的  $r$  重多分辨率分析(MRA)， $\psi = (\psi_1, \psi_2, \dots, \psi_r)^T$  是相应的多小波函数，则存在  $r \times r$  的矩阵  $H_k$  和  $G_k$ ，使得  $\phi$  和  $\psi$  满足如下双尺度方程：

$$\phi(x) = \sum_{k \in Z} H_k \phi(2x - k)$$

$$\psi(x) = \sum_{k \in Z} G_k \psi(2x - k)$$

其中， $H_k$  和  $G_k$  分别表示  $r$  重低通和高通多小波滤波器系数， $Z$  为整数。

一个三级多小波变换如图 6(b)所示，它把传感数据矩阵  $D$  每一列作为一个数据单元，多小波的重数  $r$  和  $D$  中列向量的维数  $r$  一致。在多小波变换中，每次输入的不是单个数据，而是一个列向量。这样就解决了数据集的行列不对称问题。从另一个角度来看，传感数据集  $D$  中的每一个列向量都是经过弯曲路径计算和最佳匹配点选择得到的，具有最大的空间相关性，因而特别适合多小波变换。

依据小波 Mallat 算法，由上述多小波双尺度方程可以推导出针对数据矩阵  $D$  的多小波变换公式。对第一级多小波变换，分解公式为

$$\begin{pmatrix} l_k^1 \\ l_k^2 \\ \vdots \\ l_k^r \end{pmatrix} = \sum_n H_{n-2k} \begin{pmatrix} d_n^1 \\ d_n^2 \\ \vdots \\ d_n^r \end{pmatrix}; \quad \begin{pmatrix} h_k^1 \\ h_k^2 \\ \vdots \\ h_k^r \end{pmatrix} = \sum_n G_{n-2k} \begin{pmatrix} d_n^1 \\ d_n^2 \\ \vdots \\ d_n^r \end{pmatrix}$$

重构公式为

$$\begin{pmatrix} d_n^1 \\ d_n^2 \\ \vdots \\ d_n^r \end{pmatrix} = \sum_k H_{n-2k} \begin{pmatrix} l_k^1 \\ l_k^2 \\ \vdots \\ l_k^r \end{pmatrix} + \sum_k G_{n-2k} \begin{pmatrix} h_k^1 \\ h_k^2 \\ \vdots \\ h_k^r \end{pmatrix}$$

其中， $(d_n^1 \ d_n^2 \ \dots \ d_n^r)^T$  表示矩阵  $D$  的第  $n$  列数据向量， $(l_k^1 \ l_k^2 \ \dots \ l_k^r)^T$  和  $(h_k^1 \ h_k^2 \ \dots \ h_k^r)^T$  分别表示低频和高频多小波系数， $H_{n-2k}$  和  $G_{n-2k}$  与双尺度方程一致，表示  $r$  重低通和高通多小波滤波器系数，而下标  $2k$  表示向下 2 采样。若要进行第二级多小波变换，则继续对低频多小波系数  $(l_k^1 \ l_k^2 \ \dots \ l_k^r)^T$  进行分解，如此直至第  $L$  级多小波变换。

## 4 算法分析与模拟实验

### 4.1 算法性能分析

本节从时间复杂度、计算和存储开销等方面分析最小距离弯曲路径迭代算法、最佳匹配点选择算法和多小波变换的性能。算法产生的能量开销则与压缩效率成正比，将在 4.2 节进行讨论。

弯曲路径迭代算法由初始化步和迭代步组成，在每一个迭代步比较 3 个数据点一点对的距离。因此，该算法最多需比较  $3 \max(M, N)$  次，时间复杂度

为  $O(\max(M, N))$ 。弯曲路径遵循 DTW 连续性和单调性约束条件，因而迭代步比较是邻近的 3 个点一点对。这样，算法就无需储存大小为  $M \times N$  的距离矩阵  $D\_matrix$ ，最大只需存储  $M + N - 1$  个路径元素。

最佳匹配点选择算法的时间复杂度来源于预测函数的拟合和匹配点比较。本文采用最小二乘法拟合预测函数，时间复杂度为  $O(M) + O(N)$ 。 $D_i$  和  $D_j$  存在一对多关系，即  $D_i$  中有数据点与  $D_j$  中的多个数据点对应。设这些数据点的最大数量为常数  $J$ ，则匹配点比较产生的时间复杂度为  $O(M) + O(JM)$ 。因此，总的时间复杂度为  $O(M) + O(N)$ 。另外，最佳匹配点选择算法不需要额外的存储空间。

从多小波变换的分解公式看，每计算一个多小波系数，只需要一个支撑长度的原始传感数据或同样数量的上一级低频小波系数。因此，执行多小波变换无需缓存整个二维传感数据矩阵  $D$ 。从分解和重构公式可以看出，对多小波变换需要进行矩阵运算。而对二维单小波变换而言，则在行变换后需要进行列变换，相当于矩阵运算的分解。因此，和单小波变换比较，多小波变换理论上也没有产生额外的计算量。将在模拟实验部分对此进行验证。

### 4.2 模拟实验

本节在 Matlab 环境下进行模拟实验，测试压缩算法的效率和重构精度。本文研究的重点是传感器网络中的数据压缩方法，属于应用层问题，不涉及底层协议，因此没有使用 NS-2 等网络模拟器。实验平台为 Matlab 7.0，实验环境为 PC 机，硬件配置为 Core 2 CPU (1.8 GB)，2 GB 内存。实验数据取自热带大气海洋监测项目的 15 种数据集(TAO, <http://www.pmel.noaa.gov/tao/>)。传感器网络由 100 个节点组成，随机部署在 100 m×100m 的区域内，汇聚节点接收到 2 个传感器节点的数据后进行小波变换和 8 bit 量化处理。本实验以二重 CL 多小波和 Daubechies 小波系中的 DB2 单小波为例，并以重构精度、压缩效率为评价指标，与经典的分布式单小波压缩算法(DWC)进行比较。

#### 1) 压缩效率

数据压缩效率用小波系数的能量聚集比  $ER$  和算法运行时间  $T$  表示。 $L$  级小波变换的  $ER$  定义如下

$$ER = \frac{E(A_L)}{E(A_L) + \sum_{i=1}^L E(H_i)}$$

其中， $E(A_L)$ 表示第  $L$  级低频系数的平方和， $E(H_i)$  为第  $i$  级高频系数的平方和。

实验结果如图 7~图 10 所示。图 7 和图 8 使用的测试数据是 20℃等温线的平均海洋深度。从图 7 可以看出，和 DWC 相比，本文提出的算法有更少的运行时间。图 8 表明，本文提出的算法比 DWC 有更高的低频系数能量聚集比。能量聚集比越高，则小波系数越利于编码压缩，因而会产生越好的压缩性能。在本实验中，对小波系数进行了相同的 8 bit 量化，但没有做熵编码，以利于公平比较小波变换的效率。图 9 和图 10 使用的测试数据是不同地理位置的相对湿度，其结果在趋势上与图 7 和图 8 基本一致。

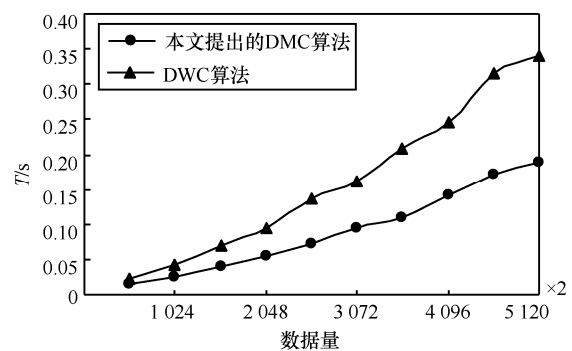


图 7 运行时间(海洋深度数据)

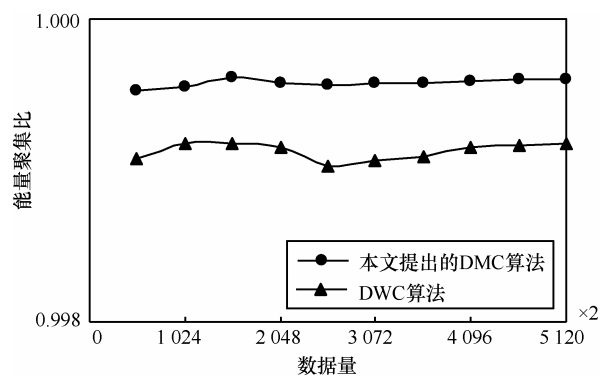


图 8 能量聚集比(海洋深度数据)

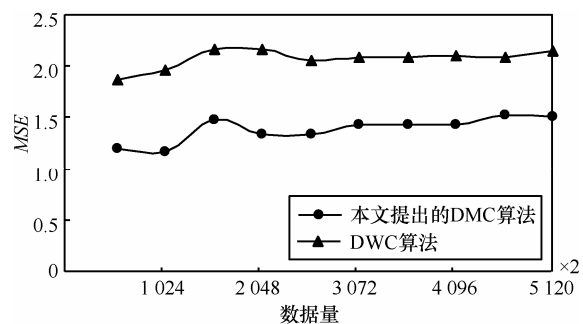


图 9 运行时间(相对湿度数据)

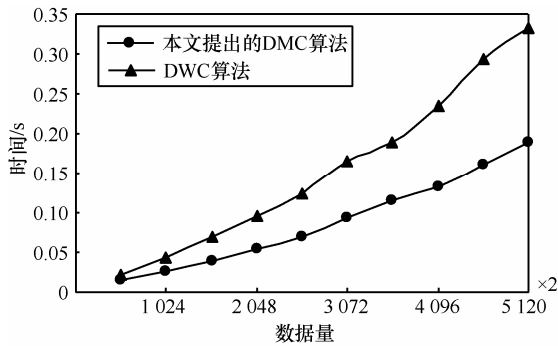


图 10 能量聚集比(相对湿度数据)

2) 重构精度

数据重构精度用原始传感数据  $d_x^y$  和重构数据  $\tilde{d}_x^y$  之间的均方差(MSE)来衡量。MSE 定义如下

$$MSE = \frac{1}{r \times R} \sum_{x=1}^R \sum_{y=1}^r (d_x^y - \tilde{d}_x^y)^2$$

实验结果如图 11 和图 12 所示。图 11 使用平均海洋深度为测试数据，对 8 bit 量化而言，本文提出的算法比 DWC 有更小的均方差，也就是说有更高的重构精度。图 12 改用相对湿度数据进行实验，对应结果与图 11 一致，表明本文的算法对不同的数据集相对稳定。

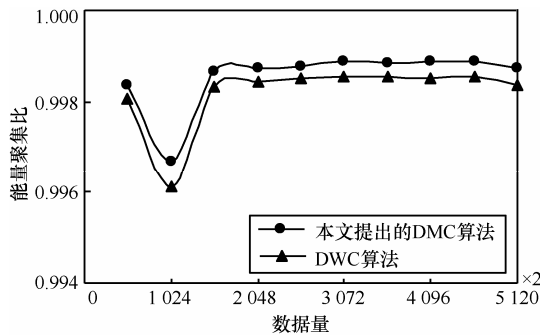


图 11 均方差(海洋深度数据)

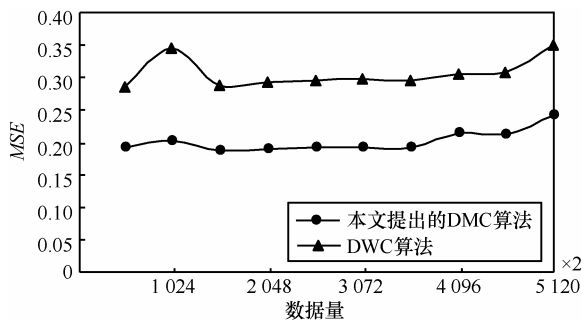


图 12 均方差(相对湿度数据)

4.3 讨论

与 DWC 算法相比，提出的 DMC 算法有一个

DTW 预处理过程，相应地增加了额外的传感器网络计算开销。虽然动态弯曲路径计算需要一定的计算量，但多小波变换每次变换两行数据，比单小波更加快捷，因此本文提出的算法在总体计算开销上仍低于 DWC 算法。这表现在图 6 和图 9 上，DMC 比 DWC 算法有更少的运行时间。

除深度和相对湿度外，TAO 项目组还提供了温度等其他 15 种数据集。对于海洋表面温度数据，DWC 反而比本文提出的算法有更高的能量聚集比。分析其原因，TAO 数据集各地点的温度数据变化平缓且相对同步，异步性不明显，因而 DWC 算法取得了好的效果。不平滑性是数据异步特性的一个表现。本文提出的算法在处理传感数据异步性方面做了大量工作，如弯曲路径迭代算法、最佳匹配点选择算法等，因此对深度和相对湿度等具有异步特性的数据取得了好的压缩效率。

在图 10 中，当数据量从 512 × 2 个数据增加到 1024 × 2 时，能量聚集比有一个下降的过程。这表明小波算法的压缩比除与算法本身有关以外，还依赖于具体的数据集，这也正是小波算法自适应性特征的表现。在 1024 × 2 这个数据点，2 个算法都出现了能量聚集比的下降，但本文的算法仍然有相对高的能量聚集比。

5 结束语

传感器网络资源有限，难以存储和传输大数据量的传感数据集。传感器网络中的数据因为节点睡眠/唤醒机制和多跳通信延迟而具有异步性特征。本文基于 DTW 技术研究传感器网络小波数据压缩问题，提出了最小距离弯曲路径迭代算法、最佳匹配点选择算法和多小波变换方法。实验结果表明，对于相对湿度、海洋深度等数据幅度跳跃比较大的数据集，本文提出的算法有更少的运行时间，更低的数据重构均方差和更高的低频系数能量聚集比。

如何依据传感数据的异步性特征选择具有合适数据性质的多小波函数，从而进一步提高压缩效率是有待深入研究的课题。

参考文献：

[1] WU C, TSENG Y. Data compression by temporal and spatial correlations in a body-area sensor network: a case study in pilates motion recognition[J]. IEEE Transactions on Mobile Computing, 2011, 10(10):

1459-1472.

[2] 尹震宇, 赵海. WSN 中基于分簇路由的多维度数据压缩算法研究[J]. 电子学报, 2009, 37(5): 1109-1114.  
YIN Z Y, ZHAO H. Research on multi-dimensional data compression algorithm for cluster-based routing in wireless sensor network[J]. Acta Electronica Sinica, 2009, 37(5):1109-1114.

[3] GANESAN D, GREENSTEIN B, ESTRIN D, *et al.* Multiresolution storage and search in sensor networks[J]. ACM Transactions on Storage, 2005, 1(3):277-315.

[4] SHEN G, ORTEGA A. Joint routing and 2D transform optimization for irregular sensor network grids using wavelet lifting[A]. Proc of IPSN[C]. St Louis, USA, 2008.

[5] NARANG S K, SHEN G, ORTEGA A. Unidirectional graph-based wavelet transforms for efficient data gathering in sensor networks[A]. Proc of IEEE International Conference on Acoustics Speech and Signal Processing (ICASSP)[C]. Dallas, USA, 2010.

[6] LIU K.H, TENG W G, CHEN M S. Dynamic wavelet synopses management over sliding windows in sensor networks[J]. IEEE Transactions on Knowledge and Data Engineering, 2010. 22(2):193-206.

[7] REIN S, REISSLEIN M. Low-memory wavelet transforms for wireless sensor networks: a tutorial[J]. IEEE Communications Surveys & Tutorials, 2011, 13(2):291-307.

[8] 周四望, 林亚平. 无线传感器网络中的小波方法[M]. 湖南: 湖南大学出版社, 2011.  
ZHOU S W, LIN Y P. Wavelet Methods for Wireless Sensor

Network[M]. Hunan: Hunan University Press, 2011.

[9] 胡玉鹏, 林亚平, 周四望等. 面向异步通信机制的无线传感器网络及其 MAC 协议研究[J]. 计算机学报, 34(8): 1463-1477, 2011.  
HU Y P, LIN Y P, ZHOU S, *et al.* Asynchronous communication mechanism oriented wireless sensor networks and MAC protocols[J]. Chinese Journal of Computers, 2011, 34(8):1463-1477.

[10] YANIV R, BURSHTAIN D. An enhanced dynamic time warping model for improved estimation of DTW parameters[J]. IEEE Transactions on Speech and Audio Processing, 2003, 11(3):216-228.

[11] HEINZELMAN W, CHANDRAKASAN A, BALAKRISHNAN H. An application-specific protocol architecture for wireless microsensor networks[J]. IEEE Transactions on Wireless Communications, 2002, 1(4):660-670.

作者简介:



周四望 (1971-), 男, 湖南岳阳人, 博士, 湖南大学副教授, 主要研究方向为传感器网络信息处理、压缩感知与小波分析。

李兰 (1984-), 女, 福建福州人, 湖南大学硕士生, 主要研究方向为传感器网络信息处理。

(上接第 85 页)

[7] MOSER A, KRUEGEL C, KIRDA E. Limits of static analysis for malware detection[A]. Proceedings of the 23rd Annual Computer Security Application Conference[C]. Seoul, Korea, 2007.421-430.

[8] BISHOP M A. The Art and Science of Computer Security[M]. Boston: Addison-Wesley Longman Publishing Co, 2002.213-217.

[9] [http://www.symantec.com/security\\_response/writeup.jspdocid=2011-022303-3344-99\[EB/OL\].2001](http://www.symantec.com/security_response/writeup.jspdocid=2011-022303-3344-99[EB/OL].2001).

作者简介:



文伟平 (1976-), 男, 湖南益阳人, 博士, 北京大学副教授, 主要研究方向为网络攻击与防范、软件安全漏洞分析、恶意代码研究、信息系统逆向工程和可信计算技术等。



梅瑞 (1984-), 男, 安徽六安人, 北京大学硕士生, 主要研究方向为网络与软件安全、信息系统风险评估等。

宁戈 (1988-), 男, 山西大同人, 北京大学硕士生, 主要研究方向为系统及网络安全、漏洞分析及利用技术等。

汪亮亮 (1987-), 男, 安徽安庆人, 北京大学硕士生, 主要研究方向为系统及网络安全、网络攻击与防范等。