

## 传感器网络基于 Voronoi 网格的数据压缩算法\*

谢志军<sup>1</sup>, 王雷<sup>2+</sup>, 陈红<sup>1</sup>

<sup>1</sup>(中国人民大学 信息学院,北京 100872)

<sup>2</sup>(湖南大学 软件学院,湖南 长沙 410082)

### Algorithm of Voronoi Tessellation Based Data Compression over Sensor Networks

XIE Zhi-Jun<sup>1</sup>, WANG Lei<sup>2+</sup>, CHEN Hong<sup>1</sup>

<sup>1</sup>(College of Information, Renmin University of China, Beijing 100872, China)

<sup>2</sup>(College of Computer, Hu'nan University, Changsha 410082, China)

+ Corresponding author: E-mail: 13308404743@hn165.com

**Xie ZJ, Wang L, Chen H. Algorithm of Voronoi tessellation based data compression over sensor networks. Journal of Software, 2009,20(4):1014–1022.** <http://www.jos.org.cn/1000-9825/3264.htm>

**Abstract:** This paper proposes a distributed multiscale data compress algorithm which can transform irregular sample data. Considering the characteristics and location information of nodes in sensor networks, a novel distributed domain partition mode DDPM (distributed domain partition model) is proposed first. On the basis of this model, a multiscale data compress model—MDCM (multiscale data compress model) is proposed for sensor networks. MDCM uses Voronoi tessellation partition the domain created by DDPM. Theoretical analyses and simulation results show that the novel methods above have good ability of approximation, and can compress the data efficiently, reduce the amount of data greatly.

**Key words:** sensor networks; Voronoi tessellation; multiscale; data compression

**摘要:** 提出了一种传感器网络中分布式多分辨率数据压缩算法.在分布式域剖分模型 DDPM(distributed domain partition model)基础上,提出了一种多分辨率的数据压缩模型 MDCM(multiscale data compress model),DDPM 把传感器网络按域划分来构建连通核,传感节点只需在连通核中寻径,因而能够明显减少寻径时间复杂度并且具有更好的分布性;MDCM 利用 Voronoi 网格来对 DDPM 所形成的域中的节点进行划分,然后采用多分辨率方法构建数据压缩模型.理论分析和实验仿真结果表明,MDCM 具有很好的逼近性能,并且能够对传感器网络中的数据进行有效压缩,可以更大程度地降低传感器网络中的数据传输量.

**关键词:** 传感器网络;Voronoi 网格;多分辨率;数据压缩

中图法分类号: TP393 文献标识码: A

\* Supported by the National Natural Science Foundation of China under Grant Nos.60673138, 60603046 (国家自然科学基金); the Natural Science Foundation of Fujian Province of China under Grant Nos.2007J0310, 2008J0012, 2008H0001(福建省自然科学基金); the Youth Foundation of Fujian Province of China under Grant No.2007F3003 (福建省青年人才项目); the Cooperation Project of Human Source Between Fujian and Hongkong of China (闽港人才合作项目)

Received 2007-02-15; Accepted 2007-11-20

集信息传感、数据处理、GPS(global positioning system)定位及网络通信功能于一体的传感器在环境与军事监控、地震与气候预测、地下、深水以及外层空间探索等许多方面都具有广泛的应用前景<sup>[1-6]</sup>,而外界环境的不确定性经常导致需要布置成百上千这样的传感器协同工作,因此,对由大规模传感器节点构成的传感器网络(sensor networks)的研究正引起研究人员的广泛关注,被认为是 21 世纪的一项挑战性研究课题<sup>[2]</sup>.

在很多传感器网络的实际应用中,传感节点是由飞机撒播的.假定飞机撒播了一定数量的能量受限的传感节点在一个区域,用这些节点来收集数据并且传送给远方的中心(sink)节点,一种简单的办法是每个节点把收集的数据进行编码量化以后直接传给中心节点.然而,这种方法忽略了一个很重要的问题,那就是空间相近的传感节点所采集到的数据具有很大的相关性.若将此数据看成一个信源,可知其信息熵非常低,也即数据的冗余度非常大,若将此数据不经处理直接发送到中心节点,将会不必要地消耗许多能量.既然数据存在很大的相关性,因此可以在传感节点对数据进行压缩后再进行传送,虽然在数据压缩的过程中会消耗掉一部分能量,但是在传感器网络中相对于通信开销而言,计算开销要小得多<sup>[6]</sup>.于是,许多研究者提出了使用基于数据压缩的数据采集算法,如文献[7-9]都采用数据压缩的方法进行数据采集.但是以上算法都假定传感节点在网络内是规则部署的,虽然规则的小波变换确实在某些场合取得了很好的效果,但是这些往往是作为有限的滤波器加以实现的,并且往往会导致多余的通信开销.在很多实际情况下,传感节点的部署是不规则的,针对规则部署的网络提出的数据转换算法往往是不实用或者应用的场合不是很多.本文对文献[2]中提出的分布式数据汇聚模型 DDAM(distributed data aggregation model)稍加改进后得到了一种分布式域剖分模型 DDPM(distributed domain partition model),然后,在 DDPM 的基础上,提出了一种应用于非规则场合的多分辨率的数据压缩模型 MDCM(multiscale data compress model);DDPM 是采用连通支配集<sup>[10]</sup>的传感器网络定向传播模型,首先提出一种域的概念,并且按域来对传感器网络进行剖分构成连通核,因而具有更好的分布性.同时,传感节点只需在连通核中进行路由,所以能够明显降低寻径的开销.MDCM 在 DDPM 的基础上提出了一种应用于非规则场合的多分辨率的小波变换模型.理论分析和仿真实验结果都表明,MDCM 具有很好的逼近能力,能够显著降低网络中传送的数据量.

## 1 相关工作

虽然近年来出现了很多优秀的小波理论,但是应用小波分析于不规则空间上的数据上是一个新的挑战.近年来,第二代小波,如提升小波(lifting scheme)<sup>[11]</sup>在很多方面已经取代第一代小波对规则空间数据进行压缩,把 lifting scheme 应用于二维的场景已被许多文献证明是一个非常大的挑战.在文献[12]中使用 non-redundant 技术把 lifting scheme 应用于二维的场景来压缩不规则的空间数据,但是这些技术都是在集中式数据库中进行操作,不能直接应用于传感器网络这种分布式环境中.例如,在文献[12]中可以自由地构造适合它们自己的多分辨率分析层次结构,但在传感器网络中,这种多分辨率的层次结构通常是由路由的层次结构决定的.近来,一些研究人员把小波变换应用于传感器网络的数据采集算法中,如在文献[7]中提出了一种基于 lifting 的分布式小波压缩算法,通过和邻居节点交换信息来消除数据的相关性.该算法是针对单跳的网络设计的,传感节点直接把数据传送到中心节点.应用这种方法的传感节点要求传送信息的距离非常远而且通信能力要求非常强,因此很难应用于实际的应用场合中.文献[8]在文献[7]的基础上提出了一种适合多跳网络的算法.在文献[13]中提出了一种分布式的多分辨率的数据采集方法,在该文中提出了两种算法,第一种算法是在当数据传送到中心节点的过程中采用 Harr 小波进行数据压缩,第 2 种采用自适应算法把传感节点分成几个组,然后在每个组里采用 Harr 小波进行数据压缩.Dimensions<sup>[9]</sup>中提出了一种网内的多分辨率小波变换和基于 wavRoute 路由的分层系数路由算法.但是,以上算法都是假定传感节点在网络内是规则部署的,虽然规则的小波变换确实在某些场合取得了很好的效果,但是这些往往是作为有限的滤波器加以实现的并且往往会导致多余的通信开销.文献[14]提出了一种多分辨率的数据分析和处理方法.

## 2 域剖分模型和基于 Voronoi 网格的多分辨率数据压缩模型

### 2.1 域剖分模型DDPM

**定义 1(核).** 设图  $G$  的节点集  $C \subseteq V$  为核,当且仅当节点集  $C$  满足以下条件: $\forall p \in V \Rightarrow p \in C$  或  $p$  为  $C$  中的某个节点  $q$  的邻节点.

**定义 2(连通核).** 给定一个图  $G=(V,E)$ ,若图  $G$  的节点集  $C \subseteq V$  为满足如下条件的节点集合:由  $C$  导出的子图是连通图,且  $C$  是图  $G$  的一个核,则称  $C$  为连通核.

**定义 3(域).** 任意传感节点  $p$ ,假定其对 Sink 节点  $R$  的相对坐标为  $(x,y)$ ,Sink 节点  $R$  及各传感节点具有相同的有效通信半径  $r$ ,则称  $p$  属于域  $(m,n)$ ,当且仅当如下公式成立: $m = \left\lceil x \frac{r}{\sqrt{2}} \right\rceil, n = \left\lceil y \frac{r}{\sqrt{2}} \right\rceil$ .其中,“/”为除法运算符;“ $\lceil$ ”为取大于等于的整数运算符.

**引理 1.** 假定 Sink 节点  $R$  及各传感节点具有相同的有效通信半径  $r$ ,则任意传感器网络存在满足定义 3 中公式的唯一域划分.

**引理 2.** 同一域中的各传感节点互为邻节点.

对任意传感节点  $P$ ,假定其地理坐标为  $(x,y)$ ;假定 Sink 节点  $R$  及各传感节点具有相同的有效通信半径  $r$  (communication radius),则任意传感节点  $P$  可以依据定义 3 计算出自己所属的“域”.同时,根据引理 1 和引理 2 的域划分规则对文献[2]中提出的分布式数据汇聚模型 DDAM 稍加改进,得到以下分布式域剖分模型 DDPM(我们规定 DDPM 操作于定义 2 给定的连通核):

分布式域剖分模型 DDPM:

- 1) 每个节点  $p$  利用 GPS 计算自己的地理坐标、剩余能量,并利用定义 3 计算出自己所属的域;
- 2) 每个节点  $p$  周期性地与其所有邻节点交换如下信息:(i)  $p$  的当前状态  $S_p$ (分为核心、网关或成员 3 种状态);(ii)  $p$  的剩余能量  $E_p$ ;(iii)  $p$  所属的域  $G_p$ ;(iv)  $p$  的地理坐标;通过此操作,每个节点可以获知自己邻节点的状态、剩余能量、所属域及到邻节点的直线距离等信息.
- 3) 初始时,以图  $G$  中 Sink 节点的状态为核心,而所有传感器节点的状态均为成员.在每个周期内,每个节点  $p$  根据自己邻节点的状态  $S_p$ 、能量  $E_p$  及所属域  $G_p$  的信息,按照如下规则来计算自身新的状态:
  - ①  $p$  所属域  $G_p$  内无核心,则  $G_p$  内的各节点依据剩余能量  $E_p$  的大小随机选举一个具有最大剩余能量的节点作为本域中的核心;
  - ② 否则,若  $p$  不为核心或网关且  $p$  与其他域中的节点相邻,则令  $p$  为网关.

由上述算法的描述可知,DDPM 是完全分布式的,每个节点只需知道其邻节点的有关信息.文献[2]给出并且证明了定理 3.

**定理 3.** 若图  $G=(V,E)$  是简单连通无向图,则由算法 DDPM 得到的节点集  $\Psi=\{p|p \text{ 为核心或网关且 } p \in V\}$  是图  $G$  的一个连通核.

### 2.2 基于 Voronoi 网格的多分辨率数据压缩模型MDCM

**定义 4(伙伴节点).** 在由 DDPM 划分的域中,给定任意的两个节点  $A$  和  $B$ ,如果  $A$  到  $B$  的通信开销比  $A$  到其他节点的通信开销小,则  $A$  和  $B$  是一对伙伴节点,且  $A$  和  $B$  不能再成为其他节点的伙伴节点.

**定义 5(Voronoi 蜂窝).** 假定  $S=\{s_1,s_2,\dots,s_n\}$  为部署在二维空间  $R^2$  中的  $n$  个传感器节点, $a_i$  为传感器节点  $s_i$  所在的位置点,则称  $Cell(s_i) = \bigcap_{j=1, j \neq i}^n \{x | dist(a_i, x) \leq dist(a_j, x), \forall x \in R^2\}$  为传感器节点  $s_i$  对应的 Voronoi 蜂窝,其中,  $dist(p,q)$  表示位置点  $p,q$  之间的 Euclidean 距离.

显然,由定义 5 可知,假定  $S=\{s_1,s_2,\dots,s_n\}$  为二维空间  $R^2$  中的  $n$  个传感器节点,则  $Cell(s_1), Cell(s_2), \dots, Cell(s_n)$  定义了域空间  $R^2$  的一个剖分(partition).

**定义 6(Voronoi 网格).** 假定  $S=\{s_1,s_2,\dots,s_n\}$  为二维空间  $R^2$  中的  $n$  个传感器节点,则称  $Cell(s_1), Cell(s_2), \dots,$

$Cell(s_n)$ 定义的域空间  $R^2$  的剖分为  $S$  生成的 Voronoi 网格.

图 1 给出了域空间  $R^2$  中由某些给定传感器节点生成的 Voronoi 网格示例.

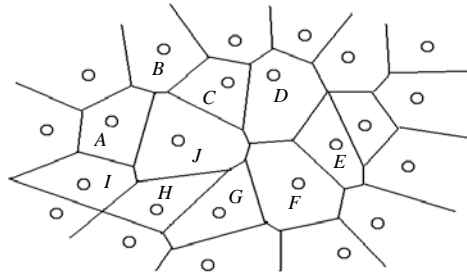


Fig.1 Voronoi tessellations created by some sensor nodes in domain  $R^2$   
图 1 域空间  $R^2$  中由某些给定传感器节点生成的 Voronoi 网格

由定理 3 可知,由 DDPM 得到的节点集  $\mathcal{V}$  构成了一个连通核,数据压缩模型必须和路由层次关系相匹配.由此,我们首先设计了一种分布式的多层次迭代路由,然后在此路由的基础上提出一个分布式的基于 Harr 小波的多分辨率数据压缩模型 MDCM.MDCM 首先用 DDPM 把网络划分成多个域,每个域对应于网络中的一个簇(cluster),然后选取簇中能量最多的传感节点作为簇头节点  $h$ ,簇内节点选取一个距离自己最近的节点作为伙伴节点,形成伙伴节点对.在每一对伙伴节点对  $(i,j)$  里,距离簇头节点较远的节点  $i$  把自身的度量值  $m$ (measurement) 传送给距离簇头节点较近的节点  $j$ ,然后在节点  $j$  计算高频系数(HP coefficient)和低频系数(LP coefficient),最后把 HP 发给簇头节点  $h$ ,而把 LP 发给更加靠近簇头的节点  $j$  进行下一级的迭代,反复迭代直到所有的 LP 最后到达簇头节点.不失一般性,如图 2(a)所示,通过 DDPM 进行域剖分后得到一个由 7 个部署不规则的传感节点组成的域,假定选定 5 号节点作为簇头  $h$ ,如节点 1 和节点 2、节点 3 和节点 4 形成伙伴节点对,迭代情况如图 2(b)所示.

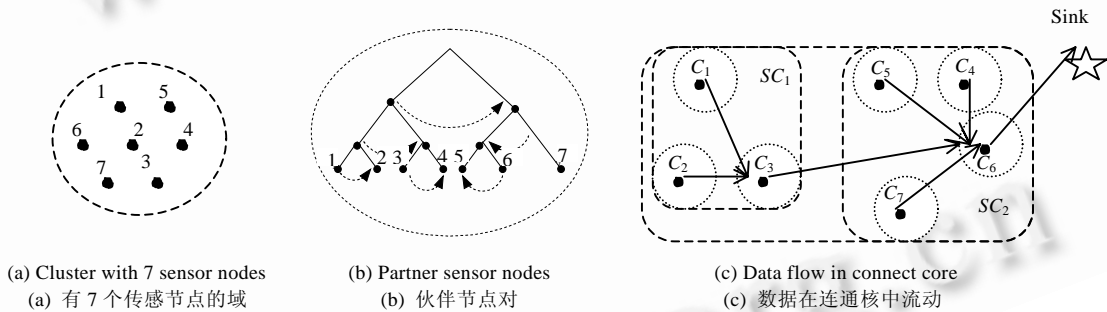


Fig.2 Data flow in connect core  
图 2 数据在连通核中流动

如上所述,单个簇的簇头节点保留所有的高频系数,但是低频系数沿着层次路由进入多分辨率压缩的下一层迭代.根据簇之间的距离和分布的地理位置,多个簇之间又组成了一个超级簇,在每个超级簇里面进行进一步的多分辨率的压缩,单个簇把已经计算出来的低频系数( $LP_i$ )传送给超级簇,超级簇根据传送来的低频系数连同本身已有的低频系数合成一个新的尺度系数(scaling coefficient)和一个小波系数集合(wavelet coefficient set),尺度系数反映了整个簇的平均性,而小波系数体现了某个度量值与平均值之间的偏差.最后,把数据沿着指定的路由传送到中心节点.如图 2(c)所示,分别由簇  $C_1, C_2, C_3$  组成了一个超级簇  $SC_1$ (super cluster 1)和簇  $C_4, C_5, C_6, C_7$  组成超级簇  $SC_2$ (super cluster 2).

MDCM 用 Voronoi 网格来对 DDPM 形成的域中的节点再次进行划分,Voronoi 网格是关于空间临近关系的一种基本数据结构,被用来划分空间的各个基本图形元素一般称为生长点.最基本的 Voronoi 网格是以平面点集

$P$  为生长点的 Voronoi 网格,  $P$  中的每个生长点  $P_i$  对应一个区域  $V_i$ , 使得  $V_i$  内的任何点与  $P_i$  的距离比其他生长点要近. 可以把 DDPM 中的传感节点看作是离散的生长点, 由这些离散的点集生成 Delaunay 三角网, 然后根据 Delaunay 三角网与 Voronoi 网格的直线对偶性质, 做每一条三角边的垂直平分线的交就构成了该点集的 Voronoi 网格, 从而间接地生成离散点集的 Voronoi 网格. 在我们的方法中, 通过首先运行 DDPM 将整个网络进行划分成多个域, 把每个域中传感节点看成生长点, 然后根据生长点来构建 Voronoi 网格(如图 3~图 6 所示). 由每个 DDPM 形成的域所生成的 Voronoi 网格可以看成是整个网络的一个局部(子)Voronoi 网格, 根据 Voronoi 网格的性质<sup>[15]</sup>, 可以将子 Voronoi 网格两两合并成大的 Voronoi 网格, 最后可以将所有的“子 Voronoi 网格”合并成一个 Voronoi 网格. 根据 Voronoi 网格的性质, Voronoi 网格中的 Voronoi tessellations 实质上是 Delaunay 三角网中的几何对偶<sup>[15]</sup>, 也即 Delaunay 三角形是 Voronoi 多边形中的几何对偶, 任意传感节点只要知道 Delaunay 的一条边就能够很容易地计算出该节点所属的 Voronoi 单元的面积.

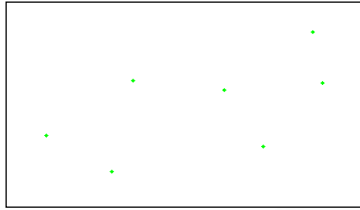


Fig.3 Set of discrete growing point

图 3 离散的生长点集

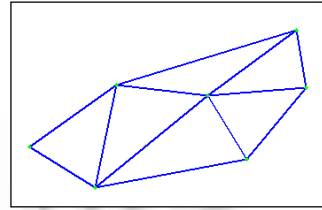


Fig.4 Delaunay triangles

图 4 生成 Delaunay 三角网

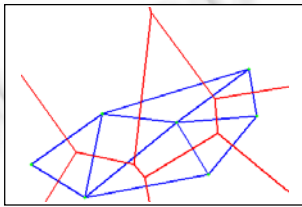


Fig.5 Voronoi tessellations created by Delaunay triangles

图 5 由 Delaunay 三角网生成 Voronoi 网格

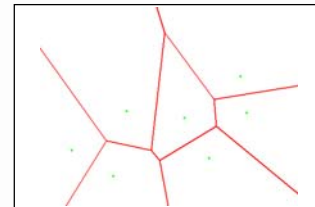


Fig.6 Voronoi tessellations

图 6 最后生成的 Voronoi 网格

给定转换模型以后, 下面我们就来详细讨论数据采集的问题. 大多数在规则环境中进行 Harr 小波数据压缩的是分段恒定信号模式(piecewise-constant signal), 在非规则环境中, 类似于文献[14,16], 我们采用一种近似分段恒定信号(approximation piecewise-constant signal)模式, 分段恒定信号是分段出现的, 且每段中的信号量相同. 在我们的模型中采用近似分段恒定信号, 可以近似地认为 Voronoi 网格里面的度量值是一个不变的常数, 用算法 DDPM 把域中传感节点划分成多个 Voronoi 网格区域, 每个传感器节点采集到的数据就能代表整个 Voronoi 网格区域的度量值. 在给定了 Voronoi 网格的面积和传感数据以后, 我们就可以把度量值及其所在的面积对应起来. 在规则环境中可以用一种非常简单的方法来计算 LP 和 HP, 对于 LP 的计算, 只是简单地对两个区域中的度量值进行求和运算, 对于 HP 的计算也只是做一下相减运算<sup>[11,13,16]</sup>; 但是, 在非规则环境中, 必须同时考虑度量值及其所对应的面积. 公式(1)和公式(2)给出近似分段恒定信号系数的计算公式.

在一个簇内, 我们根据通信代价把传感节点划分成伙伴节点对, 对于每一个伙伴节点对, 我们采用完备正交的并且能够覆盖区间  $s_1$  和  $s_2$  的基函数, 假设簇内节点对所对应的度量值分别是  $m_1$  和  $m_2$ , 则可以用下列公式(1)来计算高频系数和低频系数.

$$LP = (m_1 s_1 + m_2 s_2) \sqrt{\frac{1}{s_1 + s_2}}, \quad HP = \frac{m_1 s_1}{s_1 + s_2} \sqrt{\frac{s_2}{s_1}} - \frac{m_2 s_2}{s_1 + s_2} \sqrt{\frac{s_1}{s_2}} \quad (1)$$

单个簇内数据沿着图 2(b)中的迭代树结构进行迭代压缩运算,直到在簇头节点内产生最后一个低频系数为止.当簇内的迭代运算到达迭代树的根以后,簇头节点把低频系数连同所在簇的面积发送到超级簇内进行下一轮的压缩运算.在超级簇内我们可以得到一个低频系数和低级簇面积的集合,所以公式(1)不再适用于超级簇内的压缩运算.在公式(2)中,我们给出了多个度量值的计算高频系数和低频系数的公式:

$$LP = \frac{\sum_{i=1}^n S_i m_i}{\sqrt{\sum_{i=1}^n S_i}}, HP_i = \frac{\sqrt{S_i}}{\sum_{i=1}^n S_i} \left\{ \sum_{j=i+1}^n s_j m_j + \sum_{j=i+1}^n s_j m_i \right\} \quad (2)$$

下面我们分别给出簇内数据压缩算法和超级簇内数据压缩算法.

**Algorithm 1.** *MultiscaleCprs(M,A).*

算法输入: $M,A$  //  $M$  和  $A$  分别是度量值和对应的 Voronoi 网格的面积;

算法输出: $LP,A$ .

1. 如果度量值的个数为 1,则返回  $M$  和  $A$ ;
2. 定义一个新的度量值变量和新的面积变量: $newM \leftarrow \phi, newA \leftarrow \phi$
3. 调用配对算法计算伙伴节点对: $Match(A) \rightarrow pairs$
4. for each  $pair \in pairs$  do
5. 根据公式(2)计算  $HP$  和  $LP$
6.  $LP \rightarrow newM$
7. 计算伙伴节点的面积和: $sum(pair A) \rightarrow newA$
8. end for
9. 递归调用 *Multiscale(newM,newA)*

**Algorithm 2.** *superCompress(LPs,As).*

算法输入: $LPs,As$  //  $LPs$  和  $As$  分别是迭代树中下层  $LP$  和面积;

算法输出: $LP$  和  $HP_i$ .

1. 根据公式(3)分别计算  $LP$  和  $HP_i$
2. 返回  $LP$  和  $HP_i$

### 3 算法分析仿真

在文献[2]中已经详细论证了 DDPM 的性能,有必要讨论在 DDPM 上建立 Voronoi 网格的开销.假设一个有  $n$  个节点的传感器网络已经由 DDPM 划分成  $m$  个域,平均每个域中的节点数为  $n/m$ ;文献[17]中已经证明,对于二维空间  $R^2$  中任意给定的  $k$  个位置点,构造  $R^2$  的 Voronoi 网格的时间开销为  $O(k \log k)$ .基于上述结论,则整个网络的时间开销小于  $O(n \log(n/m))$ .

为了评价我们提出的转换方法的有效性,我们进行了一系列的仿真实验,仿真平台采用 SENSE(sensor network simulator and emulator)<sup>[17]</sup>,一个特地用来模拟传感器网络环境的平台.在 SENSE 中,被模拟的传感器网络分为 4 个层次:应用层(application)、网络层(networking)、MAC 层和物理层(physical).我们的算法是建立在应用层和网络层,在该层可以模拟 AODV(ad-hoc on demand distance vector routing)和 DSR(dynamic source routing).我们首先根据给定的有效通信距离 CR(communication radius),在一个正方形的区域里面随机地设置 3 600 个传感节点,根据我们的转换模型,我们把所有的节点设置成 5 个层级,由下到上,每个层级分别处理的簇的个数分别是 600,100,20,4 和一个最上面的根簇,这种层级的构成是随机的.在实验中,我们采用了文献[20]中的模拟数据,它们分别是:(a) 左、右两端不连续的分段恒定信号(piecewise-planar field with discontinuities in its left and right corners);(b) 二元光滑的分段信号(smooth quadratic signal);(c) 带有噪音的二元不连续分段信号(noisy quadratic sign with discontinuity),信号模式分别如图 7(a)~图 7(c)所示.图 7(d)~图 7(f)是对这 3 种信号的非线性

逼近实验结果,展示了重构误差与留存系数个数的关系.图中的曲线都展示出了一一种平滑而快速的衰减过程,这说明我们的压缩模型有很强的数据重构性能.

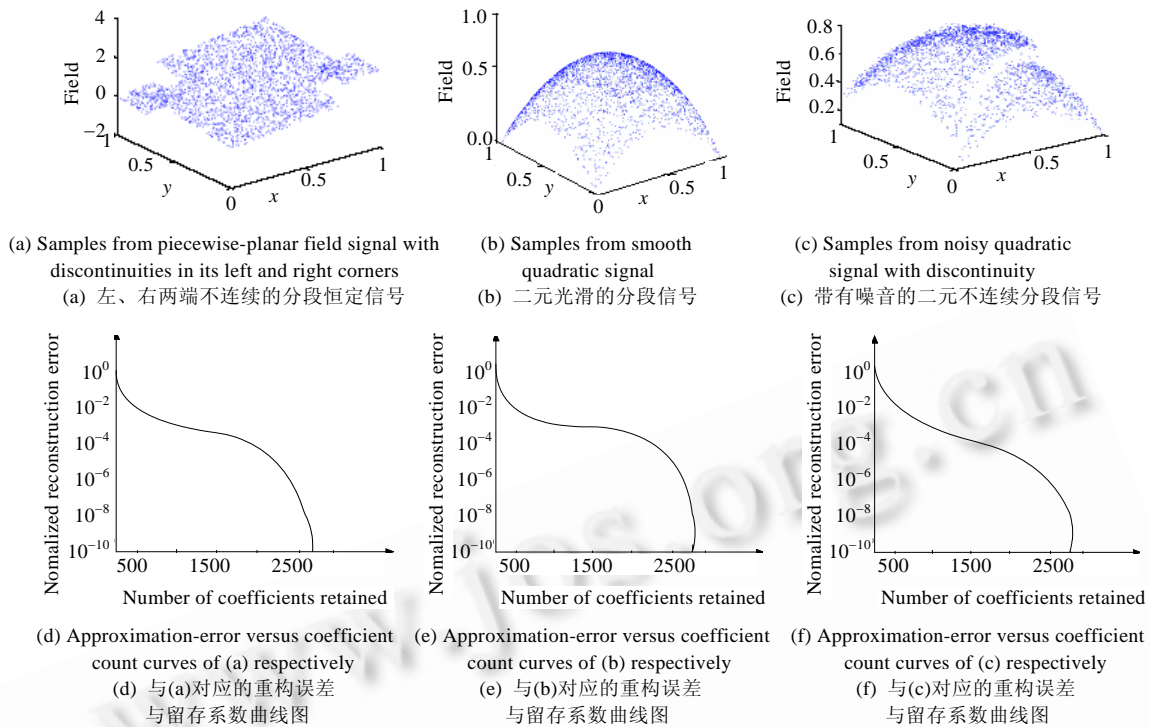


Fig.7 Approximation-Error versus coefficient

图7 重构误差与留存系数比较

为了检验在规则环境下我们的压缩算法的运行情况,我们采用了文献[14]中的实验方法.在规则环境中,将我们设计的转换算法与采用 Daubechies-2, Daubechies-4, Daubechies-6, Daubechies-8 小波进行转换的方法进行了比较.对于所有的转换,在保真度增加的情况下检验重构误差和留存系数个数的关系.在实验中,我们让传感节点随机地组合并且评价重构结果.图8表明,在规则场合下,我们提出的基于 Haar 小波的转换算法在整体上比不上 Daubechies 系列小波,但也还是很有竞争力的.这表明我们的转换算法不仅在规则领域有较强的竞争优势,而且还能很好地应用于非规则场合,而 Daubechies 系列小波则只能应用于规则场合,不能应用到非规则场合.

为了进一步验证我们所提出算法的压缩效果和压缩性能,我们采用文献[18]中的降水量数据集 (precipitation dataset) 作为我们的测试数据集. Precipitation Dataset 提供了从 1949 年~1994 年太平洋西北部地区 50km 范围内的日降水量数据.在实验中,我们采用了文献[9]中的方法,每个节点都记录了日降水量的值,然后采用第 2.2 节的数据转换和数据压缩模型进行数据转换.我们采用“压缩比(compress ratio)”作为压缩性能的判定指标<sup>[9]</sup>,第  $i$  层的压缩比是该层的原始数据量与经过处理后要向高层传送的比特数的比值.表 1 是实验结果,从表 1 我们可以看出,各层经过数据转换后得到的比特数量相差不大.这表明,我们的算法有很大的压缩性能,尤其是在高层,这一性能就越发明显,因为我们采用通信带宽作为压缩代价衡量机制<sup>[9]</sup>,层次越高就需要将更多的数据整合到某一特定的带宽中.

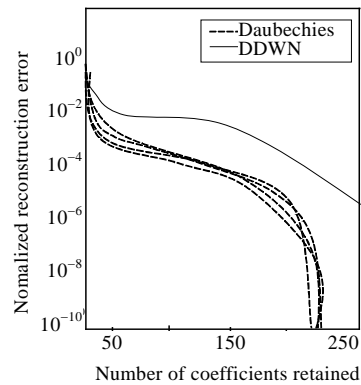


Fig.8 Results of comparing the MDCM to D-2, D-4, D-6, D-8 in a regular settings

图 8 在规则环境中 MDCM 和 D-2,D-4,D-6,D-8 的比较结果

**Table 1** Compress ratio of precipitation dataset**表 1** 压缩性能与压缩效果

Level	Raw data size (Kbits)	Mean data sent to next level (Kbits)	Compress ratio (ratio of raw data size to transmitted data size)
1	282.5	5.3	53.3
2	567.4	4.7	120.7
3	1260.5	5.2	242.4
4	5071.2	6.8	745.7
5	21179.8	8.9	2379.7

#### 4 结 语

本文提出了传感器网络中的一种完全分布式的、非规则的小波变换模型和路由。理论和实践都表明,这种模型有很强的适应性和可实现性。实验结果表明,我们设计的多分辨率的压缩技术具有很好的逼近能力和很高的压缩比,能够显著减少网络中数据的通信量,从而进一步减少了网络的开销。在以后的工作中,我们将研究比分段恒定信号更高阶信号的逼近,以求得更大的数据压缩能力。由于分布式域划分模型 DDPM 要依赖 GPS 计算自己的地理坐标,对于能量受限的传感器网络来说有一定的影响,在以后的工作中,我们将研究基能量更高效的域划分模型的多分辨率压缩算法。

#### References:

- [1] Xie ZJ, Wang L, Chen H. Subnets based distributed data-centric hierarchical ant routing for sensor networks. In: Proc. of the IEEE Int'l Conf. on Wireless Communications, Networking and Mobile Computing. Wuhan: IEEE Computer Society, 2005. 895-900.
- [2] Xie ZJ, Wang L, Lin YP, Chen H. Researches on algorithm of data aggregation based on data compression for sensor networks. Journal of Software, 2006,17(4):860-867 (in Chinese with English abstract). <http://www.jos.org.cn/1000-9825/17/860.htm>
- [3] Ren FY, Huang HN, Lin C. Wireless sensor networks. Journal of Software, 2003,14(7):1282-1291 (in Chinese with English abstract). <http://www.jos.org.cn/1000-9825/14/1282.htm>
- [4] Li JZ, Li JB, Shi SF. Concepts, issues and advance of sensor networks and data management of sensor networks. Journal of Software, 2003,14(10):1717-1727 (in Chinese with English abstract). <http://www.jos.org.cn/1000-9825/14/1717.htm>
- [5] Lin YP, Wang L. A distributed data-centric clustering hierarchical routing algorithm for sensor networks. Chinese Journal of Electronics, 2004,32(11):1883-1889 (in Chinese with English abstract).
- [6] Sun LM, Li JZ, Chen Y, Zhu HS. Wireless Sensor Networks. Beijing: Tsinghua University Press, 2005 (in Chinese).
- [7] Ciancio A, Ortega A. A distributed wavelet compression algorithm for wireless sensor networks using lifting. In: Proc. of the 2004 Int'l Conf. on Acoustics, Speech and Signal Processing—ICASSP 2004. Montreal: IEEE Computer Society, 2004. 251-254.



- [8] Ciancio A, Ortega A. A distributed wavelet compression algorithm for wireless multihop sensor networks using lifting. In: Proc. of the IEEE Int'l Conf. on Acoustic and Speech Signal Processing (ICASSP). Philadelphia: IEEE Computer Society, 2005. 825–828.
- [9] Ganesan D, Estrin D, Heidemann J. DIMENSIONS: Why do we need a new data handling architecture for sensor networks. ACM SIGCOMM Computer Communication Review, 2003,33(1):143–148.
- [10] Peng W, Lu XC. A new distributed approximate algorithm in minimum connected dominating sets. Chinese Journal of Computers, 2001, 24(3):254–258 (in Chinese with English abstract).
- [11] Sweldens W. The lifting scheme: A construction of second generation wavelets. SIAM Journal on Mathematical Analysis, 1998,29(2):511–546.
- [12] Delouille V, Jansen M, von Sachs R. Second generation wavelet methods for denoising of irregularly spaced data in two dimensions. Technical Report, Belgium: Institutde Statistique, Universit'e Catholique de Louvain, 2003.
- [13] Acimovic J, Cristescu R, Lozano B. Efficient distributed multiresolution processing for data gathering in sensor networks. In: Proc. of the IEEE Int'l Conf. on Acoustics, Speech, and Signal Processing (ICASSP), Vol.4. Pennsylvania: IEEE Computer Society, 2005. 837–840.
- [14] Wagner R, Sarvotham S, Choi H, Baraniuk R. Distributed multiscale data analysis and processing for sensor networks. Technical Report, Rice University, 2005.
- [15] Aggarwal A, Leonidas J. A linear time algorithm for computing the Voronoi diagram of a convex polygon. Discrete and Computational Geometry, 1989,4(6):591–604.
- [16] Xie ZJ, Chen H. Voronoi tessellation based multiscale data compression for sensor networks. WSEAS Trans. on Communications, 2006,10(6):779–786.
- [17] Chen G, Branch J, Pflug MJ, Zhu L, Szymanski B. SENSE: A sensor network simulator and emulator. Advances in Pervasive Computing and Networking. 2004. 249–267. <http://www.ita.cs.rpi.edu/sense/index.html>
- [18] Widmann M, Bretherton C. 50km resolution daily precipitation for the pacific Northwest. 1949–1994. <http://tao.atmos.washington.edu/data-sets/widmann>

#### 附中文参考文献:

- [2] 谢志军,王雷,林亚平,陈红.传感器网络中基于数据压缩的汇聚算法研究.软件学报,2006,17(4):860–867. <http://www.jos.org.cn/1000-9825/17/860.htm>
- [3] 任丰原,黄海宁,林闯.无线传感器网络.软件学报,2003,14(7):1282–1291. <http://www.jos.org.cn/1000-9825/14/1282.htm>
- [4] 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.软件学报,2003,14(10):1717–1727. <http://www.jos.org.cn/1000-9825/14/1717.htm>
- [5] 林亚平,王雷.传感器网络中的分布式数据汇聚层次路由算法.电子学报,2004,32(11):1883–1889.
- [6] 孙利民,李建中,陈渝,朱红松.无线传感器网络.北京:清华大学出版社,2005.27–55.
- [10] 彭伟,卢锡诚.一个新的分布式最小连通支配集近似算法.计算机学报,2001,24(3):254–258.



谢志军(1974—),男,湖南湘乡人,博士后,主要研究领域为传感器网络,数据流。



陈红(1964—),女,博士,教授,博士生导师,CCF高级会员,主要研究领域为高性能数据库,数据仓库,无线传感器网络,数据流。



王雷(1973—),男,博士后,副教授,主要研究领域为计算机网络。