第26卷 第5期
2000年9月

自 动 化 学 报
ACTA AUTOMATICA SINICA

Vol. 26, No. 5
Sep., 2000

短文

一类无纸记录仪的数据压缩研究[1)]

朱建新
(浙江大学应用数学系 杭州 310027)
(E-mail:zjx@math.zju.edu.cn)

金建祥
(浙江大学控制工程系 杭州 310027)
(E-mail:supcon@ publicl.hz.zj.cn)

摘 要 针对一类无纸记录仪数据，构造了一有效的数据压缩算法. 该算法在压缩这类无纸记录仪的历史数据中(如温度、液位、流量和压力数据)，得到了较好的应用效果，从而降低了这类无纸记录仪对内存的要求，极大地改进了它们的性能.

关键词 应用，数据压缩，无纸记录仪.

RESEARCH ON DATA COMPRESSION FOR A KIND OF NONPAPER RECORDERS

ZHU Jianxin
(*Department of Applied Mathematics, Zhejiang University, Hangzhou* 310027)
JIN Jianxiang
(*Department of Control Engineering, Zhejiang University, Hangzhou* 310027)

Abstract In this paper, an efficient compression algorithm for a kind of data for some nonpaper recorders is proposed. It is better applied to the compression of the history data, for example, temperature, liquid level, flow amount and pressure. Thus, it reduces the need of storage for these data and improves the functions of the recorders at maximum.

Key words Application, data compression, nonpaper recorder.

1 引言

近年来，随着计算机技术的发展，一种新型的记录仪——无纸记录仪[1]应运而生，它是以微处理机为核心，采用大容量半导体存储器为记录媒体，图形液晶显示器显示数字和曲线，这是传统记录仪一次最彻底的革命，完全解决了传统记录仪存在的一切问题. 但是

1) 浙江省自然科学基金资助项目(198016).
收稿日期 1998-11-19 收修改稿日期 1999-10-08

如果采样得到的数据直接进行储存，将需要大量的存储器，据估算一台8通道记录仪，按秒采样2次计算，保留10天历史数据，约需近32M SRAM 或 Flash Memory，在确保曲线复现精度的前提下，如何利用数据压缩技术，对历史数据进行有效压缩以降低对内存的要求，这是一件十分有意义的事情.

本文就如何压缩存储大量历史数据进行了探讨，提出了一种压缩数据的数学处理方法，且有较好的压缩效果.

2 数据压缩基本原理

假设采集的信号不含有阶跃型强奇异点，由于信号呈现局部振荡性，本文采用 Daubenchies 紧支撑正交基 $\{\varphi_n(2^q x-k)\}$[2]，将信号 $y(x)$ 展开成

$$y(x)\approx\sum_{k=-\infty}^{+\infty}c_k\varphi_n(2^q x-k),\qquad 0\leqslant x\leqslant m,$$

其中 $\varphi_n(x)$ 满足 $u\leqslant 0$ 或 $u\geqslant 2n-1$ 时，$\varphi_n(u)=0$. 则

$$y(x)\approx\sum_{k=2-2n}^{2^q m-1}c_k\varphi_n(2^q x-k),\qquad 0\leqslant x\leqslant m. \tag{1}$$

2.1 压缩模型

考虑信号

$$\{(j,y_j)\}_{j=0}^{m}\rightarrow\left\{\left(\frac{j}{l},y_j\right)\right\}_{j=0}^{m}$$

根据(1)式，构造

$$\hat{y}\left(\frac{x}{l}\right)\approx\sum_{k=2-2n}^{2^q m/l-1}\hat{c}_k\varphi_n\left(2^q\frac{x}{l}-k\right),\qquad 0\leqslant x\leqslant m. \tag{2}$$

可记 $\hat{l}=\dfrac{l}{2^q}$，$\hat{y}\left(\dfrac{x}{2^q\hat{l}}\right)\approx\sum\limits_{k=2-2n}^{m/\hat{l}-1}\hat{c}_k\varphi_n\left(\dfrac{x}{\hat{l}}-k\right)$，$0\leqslant x\leqslant m$. 逼近信号 $y(x)$，这里为简单起见，假定 $m/\hat{l}$ 是一个正整数.

假定 $m+3-2n>0$ 且 $\hat{l}>\dfrac{m}{m+3-2n}$，则；$m/\hat{l}-1-(2-2n)+1<m+1$；这样可实现压缩.

2.2 计算模型

记 $n_0=2-2n$，$n_1=m/\hat{l}-1$，$n_2=n_1-n_0+1$，则压缩可用如下数学模型描述：

$$\begin{cases}E=\min\limits_{\hat{c}_k\in R}\max\limits_{0\leqslant j\leqslant m}\left|y_j-\hat{y}\left(\dfrac{x_j}{l}\right)\right|=\min\limits_{\hat{c}_k\in R}\max\limits_{0\leqslant j\leqslant m}\left|y_j-\sum\limits_{k=n_0}^{n_1}\hat{c}_k\varphi_n\left(\dfrac{x_j}{\hat{l}}-k\right)\right|,\\ \text{s.t.}\quad \dfrac{m+1}{n_2}\geqslant 10,\quad x_j\in[0,m].\end{cases} \tag{3}$$

经分析计算得到用阿达姆斯最佳逼近算法[3]求解(3)式的主要步骤及参数选择的原则

1) 应避免选择 $\hat{l}(\mathrm{mod}(2))$ 是一个奇数，这样可改善数值求解的稳定性.

2) 信号的预处理. 若信号中有一些齿锯状，它使得压缩非常困难，为了解决这个问题，本文提出如下两种近似预处理方法：

i) 先滤波后压缩

根据下述公式将信号$\{y_i\}_{i=0}^m$滤波为$\{\tilde{y}_i\}_{i=0}^m$

$$\begin{cases}\tilde{y}_{i+1} = \alpha\tilde{y}_i + (1-\alpha)y_{i+1}, \\ \tilde{y}_0 = y_0,\end{cases} \quad (i = 0,1,\cdots,m-1),(0<\alpha<1),$$

然后用(2)式压缩信号$\{\tilde{y}_i\}_{i=0}^m$.

ii) 找出信号的近似包络

步骤 A. 起始点、终值点都放入最大点集、最小点集.

步骤 B. 对$\{(j,y_j),0\leqslant j\leqslant m\}$,据判断准则 $y_{i+1}-2y_i+y_{i-1}\begin{cases}>0, i \text{ 是最小点;} \\ <0, i \text{ 是最大点.}\end{cases}$找出它们的最大点和最小点.

步骤 C. 假如两个相邻最大点(或最小点)之间的信号点数大于某一给定常数 l_0,则将这些信号归入最大点集(或最小点集),其中 l_0为控制参数.

步骤 D. 据最小点集和最大点集,分别构造出三次样条插值多项式 $s_1(x)$和 $s_2(x)$;近似信号函数(加权得到):

$$y(x) \approx \omega s_1(x) + (1-\omega)s_2(x), \quad 0\leqslant\omega\leqslant 1, 0\leqslant x\leqslant m;$$

且 $\min(s_1(x),s_2(x))\leqslant y(x)\leqslant\max(s_1(x),s_2(x))$.

(注:若所有两个相邻最大点和最小点之间的信号点数足够多,则 $s_1(x_j)=s_2(x_j)=f(x_j)$,其中$(j=0,1j,\cdots,m)$,$x_j\in[0,m]$).

3) 然后用阿达姆斯算法求最大-最小值问题.

3 应用

3.1 标准曲线 $y=e^{-\frac{\ln 2}{\pi}x}\sin x$ 的数据压缩

1) $x\in[0,2\pi]$

设 $x_j=\frac{2\pi}{m}\cdot j,(j=0,1,2,\cdots,m)$,要求 $E\leqslant 0.00375$(保证相对误差限≤5‰),经计算 $n=5$时较好,计算结果如表1.

2) $x\in[0,10\pi]$

当 $x_j=10\pi\cdot\frac{j}{m},(j=0,1,2,\cdots,m),(x_j\in[0,5\times 2\pi])$时,计算结果如表2.

这些结果,表明了压缩效果与信号之间的疏密程度有关.

表1

m	$\hat{l}$	n_2	E
32	16	10	0.009 199
64	32	10	0.009 613
64	16	12	0.005 800
96	32	11	0.011 698
96	48	10	0.009 609
128	64	10	0.009 639
128	128	9	0.029 729
160	40	12	0.005 709
160	80	10	0.009 648
160	20	16	0.002 191 (相对合适)

表2

m	$\hat{l}$	n_2	E
160	20	16	0.093 104
160	16	18	0.035 359

3.2 实际数据的压缩

本文用杭州龙山化工厂的实测温度、液位、流量和压力变化数据,用模型(3)方法对它们进行了压缩,所取 $m=100$,时间间隔为10秒;然后再现压缩后的数据.

在温度数据的压缩中,如图1和图2所示,取 $n=$

$5, \hat{l}=50$，经压缩后可用 $n_2=10$ 个系数储存，压缩比为 $\frac{101}{10}=10.1$，最大偏差为1.105 755，相对误差限不超过0.27‰（≪5‰）.

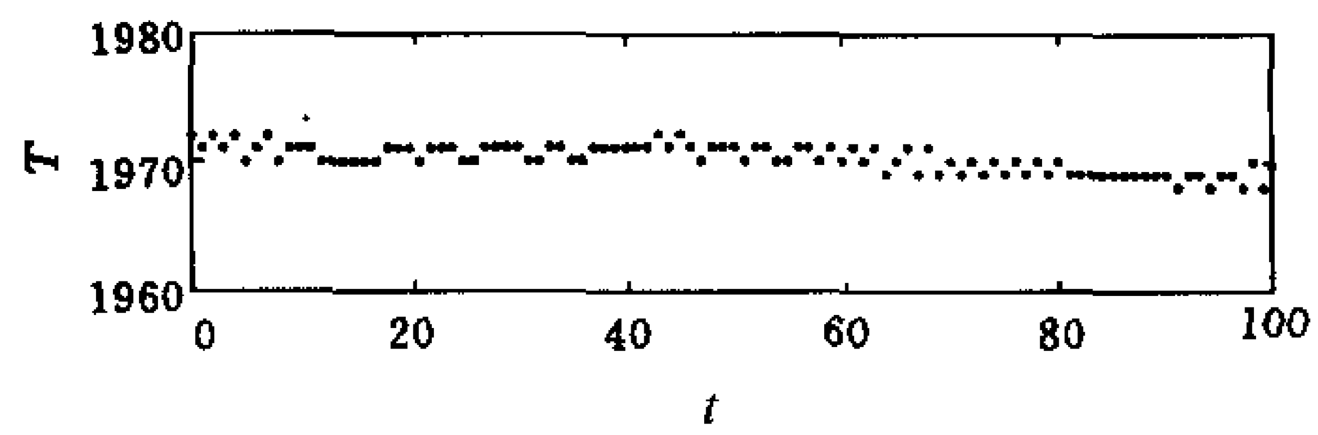

图1　原始温度数据

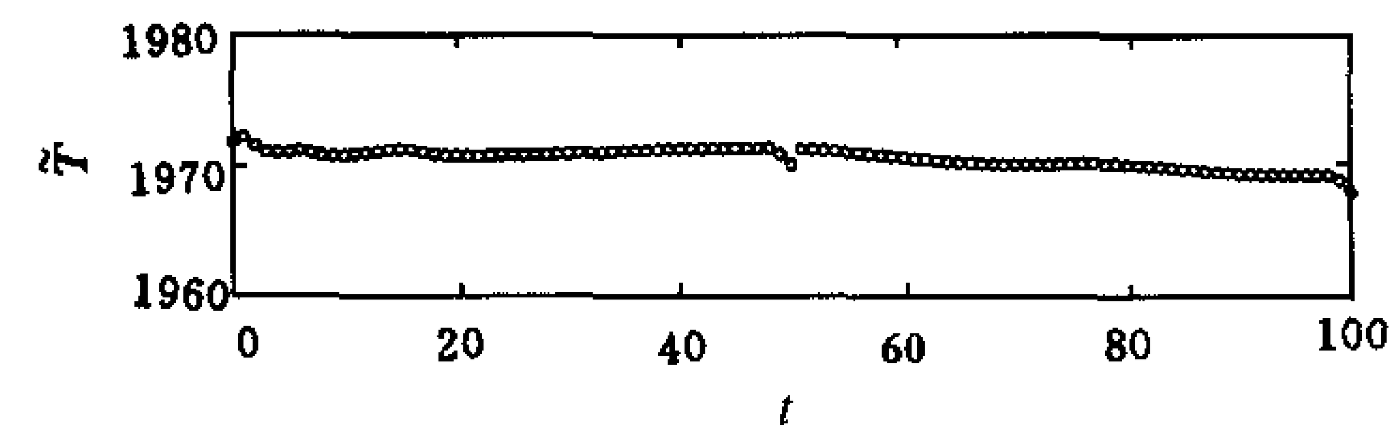

图2　压缩后的温度数据

在液位数据的压缩中，取 $n=5, \hat{l}=50$，经压缩后可用 $n_2=10$ 个系数储存，压缩比为 $\frac{101}{10}=10.1$，最大偏差为3.848 262，相对误差限不超过0.94‰（≪5‰）.

在流量数据的压缩中，取 $n=5, \hat{l}=50$，经压缩后可用 $n_2=10$ 个系数储存，压缩比为 $\frac{101}{10}=10.1$，最大偏差为14.191 399，相对误差限不超过3.46‰（<5‰）.

在变化较大的压力数据中，如图3和图4所示，取 $n=8, \hat{l}=30$，经压缩后可用 $n_2=17$ 个系数储存，压缩比为 $\frac{101}{17}\approx 6$，最大偏差为18.399 235；尽管压缩比有所减少，但是相对误差限不超过4.4‰（<5‰），满足工程上所需的精度要求.

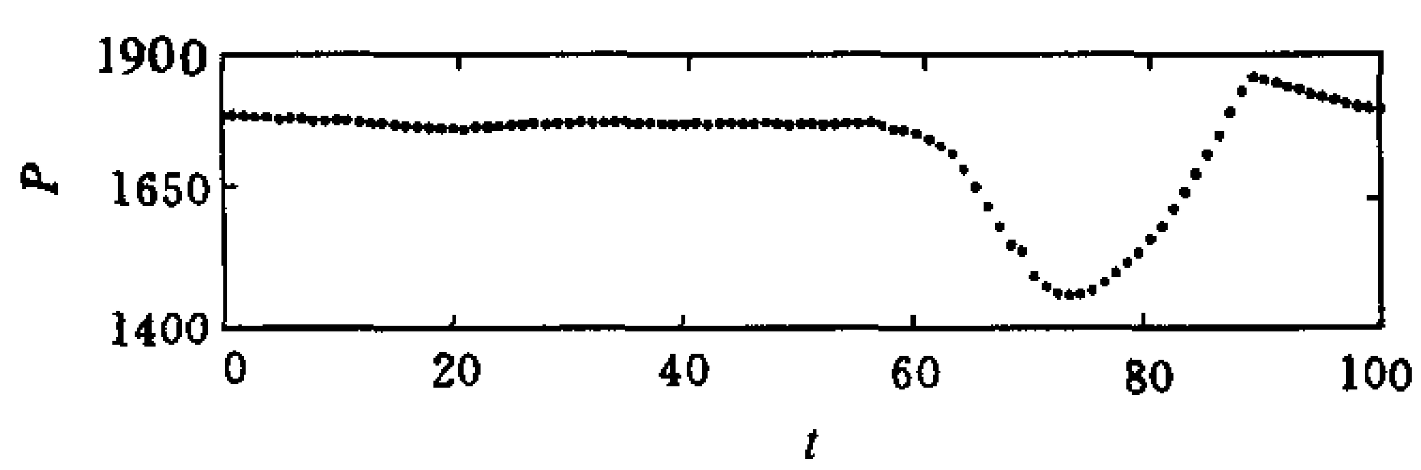

图3　原始压力数据

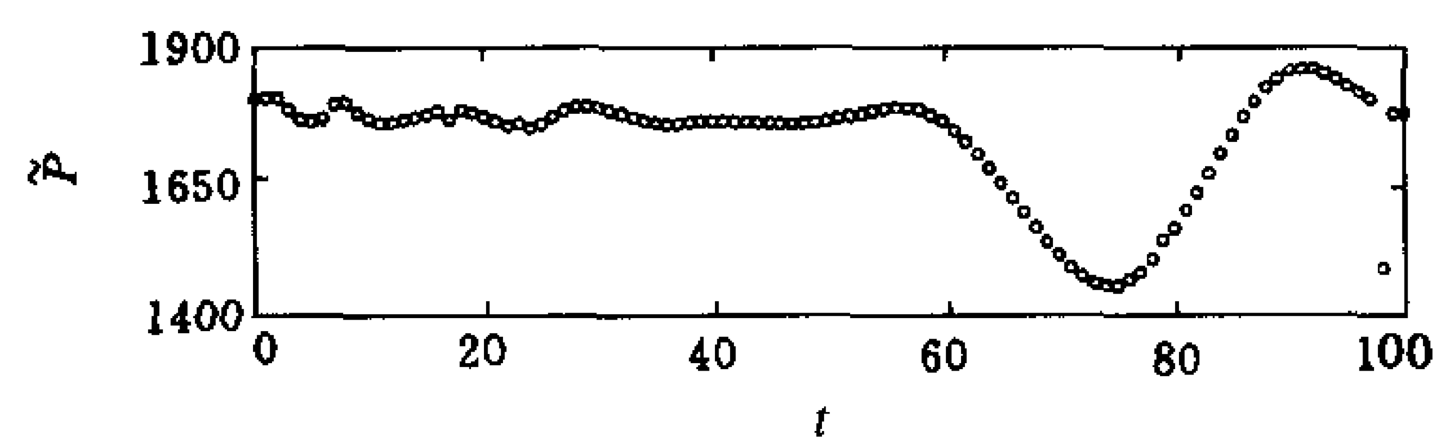

图4　压缩后的压力数据

4　结论

我们用标准曲线和实测数据（温度、液位、流量和压力）对构作的压缩计算模型(3)进行了可行性检验，结果表明：模型(3)及其实现的算法，在允许的误差范围内（相对误差限为5‰），确实有着理论价值和实用价值.运用这一方法压缩历史数据，必将极大地改善无纸记录仪的性能.

参　考　文　献

1　金建祥等. 九十年代显示记录新概念——无纸记录仪. 检测与控制装置，1995，**22**(5)：42～44

2　Ingrid Daubechies. Ten Lectures on Wavelets. Philadelphia：Capital City Press，1992

3　沈燮昌. 多项式最佳逼近的实现. 上海：上海科学技术出版社，1984

4　Murtaza Ali，Ahmed H. Twefik. Real time implementation of second generation of audio multilevel information coding. *SPIE* (*Wavelet Applications*)，1994，2242：212～223

5　Ciarlet P G. 矩阵数值分析和优化. 北京：高等教育出版社，1990

6　Yan Zhuang，John S Baras. Optimal wavelet basis selection for signal representation. *SPIE*(*Wavelet Applications*)，1994，2242：200～211

朱建新 男，副教授. 分别于1984年、1991年、1998年于浙江大学应用数学系获理学学士学位、硕士学位、博士学位. 目前研究方向为应用数学，侧重于计算逼近.

金建祥 男，副教授. 1984年7月毕业于浙江大学化工自动化专业. 主要从事DCS、自动化仪器仪表开发研究工作.

《自动化学报》征稿简则

一、《自动化学报》是中国自动化学会主办的高级学术期刊，每年出版六期.

二、本刊刊载自动化科学与技术领域的高水平学术论文和科学研究成果. 内容包括：1. 自动控制理论；2. 系统理论与系统工程；3. 自动化技术及其在国民经济各领域中的创造性应用；4. 自动化系统计算机辅助技术；5. 机器人与自动化；6. 人工智能与智能控制；7. 自动控制系统中的新概念、新原理、新方法、新设计；8. 信息理论与信息处理技术，模式识别；9. 自动化学科领域的其它重要问题.

三、本刊发表的文章以论文和短文两种形式为主，并不定期地发表综述与评论性文章、长论文、书刊评论、问题讨论、读者来信和国内外学术活动信息等.

四、本刊原则上只发表原始性稿件，但不排除刊登已在国内外学术会议上发表或准备发表的优秀论文的可能性(对于此种情况，作者必须如实说明).

五、作者投稿时需签署"作者承诺".

六、来稿格式及要求

1. 来稿要求论点明确、论证充分、语言通顺、文字简练. 一般定稿时论文不超过6000字；短文不超过3000字；其它形式文章视具体内容由编辑部决定. 对重要成果进行系统、完整叙述的长论文字数可稍长. 长论文稿件，作者在投稿时必须注明.

2. 稿件首页应包括下列内容：标题；作者姓名、工作单位、详细通讯地址(包括邮政编码)、E－mail、电话号码；论文摘要；关键词；用英文书写的上述内容.

3. 论文和短文的文章结构请参照本刊最近发表的文章格式. 论文摘要在200字以内；短文100字左右. 文中缩写词(中文或英文)须在首次出现时注明全称；公式、图、表均须分别用阿拉伯数字全文统一编号.

4. 计量单位一律用国际单位，即SI单位制. 名词术语必须规范化、标准化，前后一致. 外国人名、地名、书刊名称除已通用者外一律用原文.

5. 参考文献按文中出现的先后次序排列，文献如为期刊时，按编号，作者(姓在前如 Wiener L N, Kalman R E, Wang H.). 文章题目，期刊名(外文可根据国际惯例使用缩写词)，年份，卷号(期号)：页码顺序编排. 文献如为图书时，则按编号，作者(姓在前). 书名，版次(初版不写)，出版地点：出版者，年份，页码顺序排列. 文中未引用的文献不得列入参考文献栏目.

6. 来稿请用A4纸1.5倍行距打印.

七、作者必须对稿件内容的真实性和可靠性负责.

八、本编辑部在收稿后一周内通知作者，并在稿件修订过程中与作者保持联系. 如果作者在来稿中不作特殊说明，编辑部将只与第一作者联系.

九、已被本刊接受发表的稿件，按审查意见和"作者加工稿件须知"修改后一式两份同软盘一起寄编辑部. 并需附所有作者的简介.

十、来稿刊登与否由编委会最后审定. 编委会有权对来稿作适当文字删改或退请作者修改. 来稿一经发表，按篇酌付稿费，并赠送当期杂志1本及30份抽印本. 经审查后不拟刊登的文稿，一般情况在半年内退还.

十一、来稿(一式叁份)请寄北京市中关村中国科学院自动化研究所转《自动化学报》编辑部，邮政编码100080. E-mail：aas@ iamail. ia. ac. cn