

# K-means 聚类-DCT 压缩算法在振动传感器中的应用\*

王昱钦<sup>1</sup>, 王鑫<sup>2</sup>, 刘保强<sup>1</sup>, 李轶<sup>1</sup>, 洪晟<sup>3</sup>

(1. 江苏自动化研究所, 江苏 连云港 222000; 2. 中国人民解放军 32381 部队, 北京 100000;

3. 北京航空航天大学 网络空间安全学院, 北京 100191)

**摘要:** 为延长无线振动传感器在需要对大量高频振动数据进行采集情况下的使用寿命, 首先对现有振动数据压缩算法进行了研究, 对其存在的问题进行了分析, 并在此基础上提出了一种基于 K-means 聚类-DCT 双重数据压缩算法的压缩方法。所采用的 K-means 聚类-DCT 双重数据压缩算法针对预测性维护数据特点, 首先利用 K-means 算法对振动数据进行聚合分类, 再根据振动信号频域特点进行离散余弦变换 (Discret Cosine Transform, DCT) 压缩。通过实际验证表明, 采用该算法对振动数据进行数据聚合, 可以非常明显地提高数据压缩的效率, 从而在传输时将冗余数据有效减少。而在相同数据量情况下, 采用改进的双重压缩算法的峰值信噪比与其他算法相比, 其具有更好的性能。

**关键词:** 传感器; 低频振动; 中频振动; 高频振动; k-means 聚类-DCT 压缩

**中图分类号:** TP212

**文献标志码:** A

**DOI:** 10.16157/j.issn.0258-7998.222844

**中文引用格式:** 王昱钦, 王鑫, 刘保强, 等. K-means 聚类-DCT 压缩算法在振动传感器中的应用[J]. 电子技术应用, 2023, 49(1): 81-85.

**英文引用格式:** Wang Yuqin, Wang Xin, Liu Baoqiang, et al. Research and application of K-means clustering and DCT compression algorithm in vibration sensor[J]. Application of Electronic Technique, 2023, 49(1): 81-85.

## Research and application of K-means clustering and DCT compression algorithm in vibration sensor

Wang Yuqin<sup>1</sup>, Wang Xin<sup>2</sup>, Liu Baoqiang<sup>1</sup>, Li Yi<sup>1</sup>, Hong Sheng<sup>3</sup>

(1. Jiangsu Automation Research Institute, Lianyungang 222000, China;

2. Unit 32381 of the Chinese People's Liberation Army, Beijing 100000, China;

3. School of Cyber Science and Technology, Beihang University, Beijing 100191, China)

**Abstract:** In order to prolong the service life of wireless vibration sensors when collecting a large number of high-frequency vibration data, this paper studied the existing vibration data compression algorithms, put forward and analyzed the existing problems, and on this basis, proposed an effective mechanism of K-means clustering-discrete cosine transform (DCT) dual data compression. According to the characteristics of predictive maintenance data, K-means clustering-DCT dual compression algorithm firstly used K-means algorithm to aggregate and classify vibration data, and then carried out DCT compression according to the frequency domain characteristics of vibration signals. The verification results showed that the algorithm significantly improved the data compression efficiency and reduced the transmission of redundant data by aggregating vibration data. In addition, under the condition of the same amount of data, the algorithm had better application performance after improving the peak signal-to-noise ratio compared with other algorithms.

**Key words:** sensor; low frequency vibration; intermediate frequency vibration; high frequency vibration; K-means clustering and DCT compression algorithm

## 0 引言

在工作时机械设备基本都存在比较明显的振动情

况, 如果想要监测设备的运行状态, 是可以通过分析机械设备所产生的振动信号来实现的, 并且通过分析监测结果的变化情况可以在早期及时地发现机械设备可能

\* 基金项目: 工信部高质量发展专项 (TC210804R-2)

出现的故障情况<sup>[1]</sup>。但是如果要想通过设备所产生的振动信号对其运行状态情况获得非常准确的评估或预测结果,则需要对机械设备所产生的大量高频振动数据进行有效分析<sup>[2]</sup>,为了能够有效采集到大量高频振动数据,需要安装大量的传感器用于采集机械设备在工作时所产生的高频振动数据<sup>[3]</sup>。但是现有的基于 ZigBee、蓝牙技术等低功耗、低速率的无线传输技术是无法满足此类应用需求的,无法满足需求的原因主要有两个方面<sup>[4]</sup>: (1)对高频振动数据进行采集会快速消耗终端节点所使用的电池电量,进而影响整个采集系统的工作寿命;(2)大量数据的无线传输会严重占用系统带宽资源,从而会造成传输过程丢包率的升高,最终会影响到数据分析的准确性<sup>[5]</sup>。

对数据进行压缩的主要目的就是想要通过传感器在边缘侧的运算,较大幅度地减少数据在无线传输过程中所消耗的能耗,虽然这样会损失传感器的部分能耗<sup>[6]</sup>,但只要确保传感器所消耗的这部分能耗要大大低于传输过程所需能耗,就可以实现总能耗的减少,但并不是所有的数据压缩算法都可以达到这一目的<sup>[1]</sup>。针对上述问题,本文首先对现有多种振动数据压缩算法进行了研究,对现有压缩算法存在的问题进行了分析,并在此基础上提出了一种面向无线通信的数据压缩算法——K-means 聚类-DCT 双重压缩算法。通过实际验证表明,采用该算法对振动数据进行数据聚合,可以非常明显地提高数据压缩的效率,从而在传输时将冗余数据有效减少<sup>[7]</sup>。而在相同数据量情况下,采用改进算法的峰值信噪比与其他算法相比,具有更好的性能。

## 1 振动数据压缩研究概述

设备预测性维护是通过采集机械设备运行时所产生的振动数据,然后对振动数据进行分析,通过分析所采集到振动数据并根据分析结果来确认机械设备的运行状态是否正常。奈奎斯特采样定理指出,如果需要对模拟数字信号进行转换,在其转换的过程当中,如果采样频率是所采样信号中最高频率的 2 倍及以上时,完成采样后所得到的数字信号就能够将原始信号中的全部信息完整地保留下来<sup>[8]</sup>。机械设备可能在特定情况下产生高频振动,为防止振动数据在采样后频域出现混叠情况,在对振动数据进行采样时,有必要采用较高的采样频率。这造成了传感器将采集大量的振动数据,由于无线传感器对于功耗的要求,将大量数据直接上传至采集基站无疑将大大降低无线传感器的实际使用寿命。为有效延长传感器实际使用寿命,在确保较高采样频率的前提下,通常会应用数据压缩算法对传感器采集到的振动数据进行压缩,从而能够将传感器传输数据的效率大幅提高,并且可以将数据传输的时间大大降低,最终大幅降低传感器的实际功耗。通过压缩算法来提高数据

传输效率是振动监测领域通用做法<sup>[9]</sup>。

无线传感器所采用的压缩算法可分为无损压缩算法和有损压缩算法。其中无损压缩算法是指接收到被压缩的数据后能够将压缩后的数据完整恢复的算法,最为典型的无损压缩算法有 LZW 压缩算法、Huffman 编码、行程编码方法等<sup>[10]</sup>。LZW 压缩算法是一种基于字符字典库的无损压缩算法,该算法通过建立一个字符字典库,用一个数字来代表放入字符字典库中的第一次出现的字符串,如果这个字符串再次出现时,即可用表示它的数字来代替,然后使用较短长度编码来对较长的字符串进行标识,通过这样的方法来最终实现对数据的压缩<sup>[11]</sup>;Huffman 编码是一种用于无损压缩的熵编码算法,该算法首先会统计出需要压缩的数据当中所有字符出现的概率,然后在对数据进行重构的过程中是按照字符出现概率的高低进行的,使用较短的编码字符来代替所有出现概率高的字符,最终将整体数据量压缩到最小<sup>[12]</sup>;行程编码方法也叫 RLE 压缩编码,是 Golomb 于 1966 年提出的一种相对简单的编码技术,其原理是将由相同数值组成的连续数值串通过数值串的长度以及代表数值来表示,从而使得压缩后的数据量要小于初始的数据量<sup>[13]</sup>。目前有损压缩算法主要都是对现有算法的改进或者裁剪,其中比较典型的有损压缩算法主要有小波变换方法、时间顺序近似方法等。

对于预测性维护,在满足系统要求的前提下,可对振动数据进行有损压缩。Liu 和 Cheng<sup>[14]</sup>提出了一种应用于设备状态监测的传感器数据压缩方法,方法是用 DWT 和 DSC 实现对振动数据的压缩处理。Steffen 等人在文献[15]中提出了一种传感器的自适应压缩算法,该方案对传感器性能要求非常高,而传感器往往出于节能考虑,采用的控制处理器性能都比较低,复杂的压缩算法增加了传感器的处理时间,导致数据无法及时发送<sup>[16]</sup>。Yoon<sup>[17]</sup>提出了一种基于无线传感器网络自适应压缩率的选择方案,该方案可以根据电池的电量状况来选择不同的压缩模式,在电池电量充足时,自适应选择使用高性能的压缩方案;反之,在电池电量较低时,则自适应选择计算量相对较低的压缩方案,通过该方式实现能量使用的均衡。在文献[18]中,作者提出基于近似多项式的数据压缩算法,并将其应用于非周期信号测量中。在文献[19]中,作者提出了一种应用于无线传感器网络数据传输的压缩效率较高的无损压缩算法,该算法的核心采用了 LZW 算法<sup>[11]</sup>,实际由于 LZW 算法较复杂,对于传感器硬件压力过大,且对 RAM 要求过高,使用该算法将极大地影响传感器功耗,由于 LZW 算法为通用压缩算法,实际用于传感器数据压缩性能提高有限。

但是大部分的以上工作并不适用于机械振动的振动数据压缩,由于算法复杂度以及低功耗无线传感器的资源限制,其中有些算法不适用于此类应用场景。鉴于此,本文着重针对机械振动数据的监测,提出了一种新的高效数据压缩算法,在重复考虑算法复杂度和计算复杂度的同时,对振动数据处理有着更高的压缩效率,适用于预测性维护的应用场景。

## 2 K-means 聚类-DCT 双重压缩算法研究

没有一种压缩算法能适用于各种传感器数据的压缩,例如振动旋转电机的振动数据压缩算法,用于往返振动所采集的信号时,可能性能也存在问题。因此,算法需要针对所监测对象进行设计。机械振动信号在时域变换较快,但是在频域表现为重叠和噪声<sup>[20]</sup>。K-means 聚类-DCT 双重压缩算法针对预测性维护数据特点,本文首先利用 K-means 算法对振动数据进行聚合分类,再根据振动信号频域特点进行离散余弦变换(Discret Cosine Transform, DCT)压缩。压缩流程如图 1 所示。

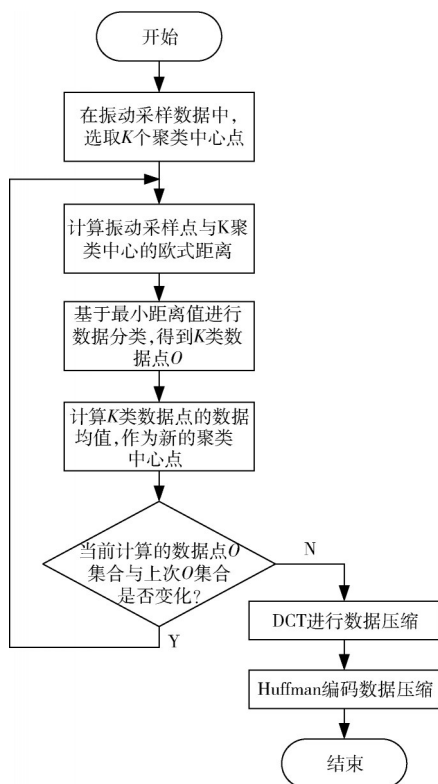


图1 数据压缩流程

针对振动数据特点对 ADC 所采集到的数据进行初步分类,主要步骤如下:

(1)将 ADC 采集的振动数据  $D = \{x_1, x_2, \dots, x_n\}$ , 任取  $k$  个数作为聚类中心,用  $M = \{m_1, m_2, \dots, m_k\}$  表示。

(2)算出  $N = \{n_1, n_2, \dots, n_m\}$  与聚类中心  $M$  之间的

欧式距离  $d(m_i, n_j) = \sqrt{\sum_{k=1}^K (m_{ik} - n_{jk})^2}$ ,  $N \in D$  且  $N \notin M$ 。

(3) 比较距离值, 依据  $d(m_i, n_i) = \min\{d(m_1, n_i), d(m_2, n_i), \dots, d(m_k, n_i)\}$ , 将振动数据  $D$  根据最小距离划分为  $k$  类  $O = \{o_1, o_2, \dots, o_k\}$ 。

(4) 重新计算  $O$  中数据均值  $p_j = \frac{1}{n_j} \sum_{i=1}^{n_j} x_i^{(j)}$  ( $j = 1, 2, \dots, k$ ), 作为新的中心  $M^*$ , 重复步骤(1)~(3)。将数据重新归类为新的  $k$  类  $O^* = \{o_1^*, o_2^*, \dots, o_k^*\}$ 。

(5) 比较  $O$  与  $O^*$  中数据情况, 如果数据依然存在变化, 重复步骤(4), 直到每类中数据对象在误差平方和函

数  $J_w = \sum_{j=1}^k \sum_{i=1}^{n_j} \|m_i^{(j)} - p_j\|^2$  收敛时, 停止迭代。

对振动信号进行 FFT 能直观地反映信号的频域特点, 振动信号可能包含多个高频域振动信号, 或是分布在宽频的随机信号构成。振动信号在基于 K-means 聚类算法处理后, 利用离散余弦变换的变换域压缩, 去除信号的相关性, 减少冗余数据。由于小波变换具有良好的时域定位和能量集中特性, 对归类后的振动信号  $f(x)$  进行 DCT 变换, 用下列公式表示:

$$F(u) = \sum_{x=0}^{N-1} f(x) k(m, x) \quad (1)$$

其中,  $k(m, x)$  为 DCT 变换的变换核。

将振动数据变换到频域后, 对变换后的数据进行量化, 并采用 Huffman 编码对变换后的数据进行二次数据压缩。

## 3 实验数据分析

无线振动传感器硬件基于 STMicroelectronics 公司低功耗单片机 STM321476 实现, 振动 MEMS 芯片采用 ADI 公司 ADXL1002, 线性度为  $\pm 0.1\%$ , 量程为  $\pm 50$  g。采样频率 20 kHz, 采用 ADI 公司 ADS8689 16 位 ADC。采用 K-means 聚类-DCT 双重压缩算法, 与标准 DCT、DWT、LZO 编码进行压缩对比。

STM321476 MCU 一款基于高性能的 Arm® Cortex®-M4 32 位内核, 最大工作频率为 80 MHz。考虑到采集的振动数据量, 传感器采用 UWB 模块实现无线通信, 将数据采集数据通过 UWB 模块发送至实验计算机, 再使用仿真工具对数据进行处理。测试环境由 SD&PTI PE 型通用振动台功放设备及泰克 AFG1062 任意波形函数信号发生器构成, AFG1062 生成正弦振荡波形, 经由功放设备放大, 驱动振动台振动。传感器硬件如图 2 所示。在真实测试环境中, 传感器采用磁吸方式吸附在传动机构电机的轴向位置, 与电机轴水平位置垂直, 并采用粘胶固定。





图2 传感器硬件图

### 3.1 算法效果评估指标

为评估算法效果,文中使用压缩率衡量算法压缩情况,压缩率定义如式(1)所示:

$$\text{压缩率} = \frac{\text{原始数据大小}}{\text{压缩后大小}} \quad (2)$$

由于DCT是有损压缩方式,本文用峰值信噪比(Peak Signal to Noise Ratio, PSNR)来评估压缩质量,计算公式如式(2)、式(3)所示。

$$\text{PSNR} = 20 \log_{10} \frac{255}{\text{MSE}} \quad (3)$$

$$\text{MSE} = \sqrt{\frac{1}{N_s} \sum_{i=0}^{N_s-1} (x_i - \hat{x}_i)^2} \quad (4)$$

其中,  $N_s$  为总采样数据量,  $\hat{x}$  为压缩后的振动数据。

### 3.2 数据分析

实验分别采集1 k、2 k、4 k、16 k个点,每个点为16 bit浮点数,每次采集10次,使用MATLAB进程处理,并对每种算法的实验结果进行均值处理,结果如表1所示。

从图3和图4的实验结果看,采用K-means算法对振动数据进行数据聚合,可以非常明显地提高数据压缩的效率,从而在传输时将冗余数据有效减少。而在相同数据量情况下,改进算法在峰值信噪比方面与其他算法相比具有更好的性能。

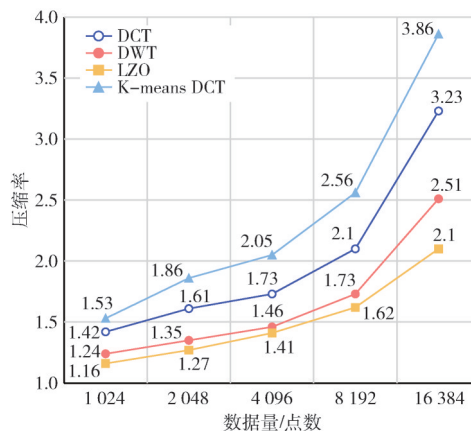


图3 压缩率曲线

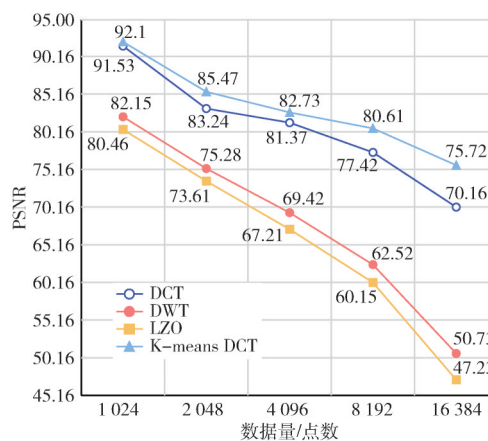


图4 PSNR曲线

## 4 结论

本文针对预测性维护振动数据的特点,对DCT算法进行了针对性的改进,并且通过K-means聚类算法对振动数据进行了进一步处理,根据数据特性将数据进行了相应分类,从而可以更好地利用DCT算法的压缩性能。通过对实验数据的处理分析,与DCT、DWT、LZO算法进行对比,双重压缩算法在压缩率和峰值信噪比上具有更好的性能。

### 参考文献

- [1] 刘锁,何青,石志刚.振动信号无线传输压缩算法的能效分析[J].电力与能源,2015,36(6):818-821.
- [2] 冯俊涛,戚厚军,谭鹏,等.基于LabVIEW的振动信号

表1 实验数据表

数据量(16 bit)	DCT算法		DWT		LZO		K-means DCT 双重压缩算法	
	压缩率	PSNR	压缩率	PSNR	压缩率	PSNR	压缩率	PSNR
1 024	1.42	91.53	1.24	82.15	1.16	80.46	1.53	92.1
2 048	1.61	83.24	1.35	75.28	1.27	73.61	1.86	85.47
4 096	1.73	81.37	1.46	69.42	1.41	67.21	2.05	82.73
8 192	2.1	77.42	1.73	62.52	1.62	60.15	2.56	80.61
16 384	3.23	70.16	2.51	50.73	2.1	47.23	3.86	75.72

- 分析与报警系统的设计[J]. 机械研究与应用, 2020, 33(3):145-148,152.
- [3] 孔祥旭. 基于 IEEE 802.11 协议的机械振动无线传感器网络高效传输方法研究[D]. 重庆: 重庆大学, 2019.
- [4] 罗巧, 刘红. 基于 Zigbee 技术的风场监控系统构建[J]. 信息网络安全, 2013(4): 36-38.
- [5] ALSALAET J K, NAJEM S I, ALI A A. Vibration data compression in wireless sensors network[C]// 2012 IEEE International Conference on Signal Processing, Communication and Computing (ICSPCC 2012), 2022.
- [6] 张永棠. 一种基于安全区域的 WSN 流量分析聚合方法[J]. 信息网络安全, 2021, 21(9): 80-89.
- [7] 朱小错. 面向结构健康监测的无线传感器网络的研究与设计[D]. 杭州: 浙江理工大学, 2010.
- [8] 王玉华, 黄凯明. 基于 LabVIEW 的冲击力检测系统设计[J]. 计算机测量与控制, 2018, 26(4): 19-23.
- [9] 刘京玮, 赵帅兵, 田乐, 等. 基于 ROS 改进算法的无线传感器网络耗能研究[J]. 信息网络安全, 2014(7): 65-68.
- [10] 刘向宇, 王雅哲, 杨晓春. 面向无线传感器网络的流数据压缩技术[J]. 计算机科学, 2007, 34(2): 141-143.
- [11] 林锋, 黄可生, 徐小杰. LZW 压缩算法在多通道高速数据采集系统中的实现[J]. 船舶工程, 2011, 33(2): 232-235.
- [12] 张宇. 面向无线传输的数据压缩算法[D]. 秦皇岛: 燕山大学, 2016.
- [13] 高艺, 孙佳玲, 李维祥, 等. 基于预测编码的无线传感器网络节点数据压缩算法研究[J]. 南开大学学报(自然科学版), 2009, 43(6): 26-30.
- [14] LIU S, CHENG L. Efficient data compression in wireless sensor networks for civil infrastructure health monitoring[C]// 2006 3rd Annual IEEE Communications Society on Sensor and Ad Hoc Communications and Networks, 2006.
- [15] STEFFEN B. Adaptive compression of large vectors [J]. arXiv:1506.00222, 2015.
- [16] 吴佳杰, 朱荣钊, 钟志峰. 无线传感器网络基于时空差分阈值 Huffman 数据压缩[J]. 计算机测量与控制, 2019, 27(2): 273-275.
- [17] YOON I, YI J M, JEONG S. Adaptive sensing and compression rate selection scheme for energy-harvesting wireless sensor networks[J]. International Journal of Distributed Sensor Networks, 2017, 13(6): 106-109.
- [18] CHAN J C, TSE P W. A novel, fast, reliable data transmission algorithm for wireless machine health monitoring[J]. IEEE Trans. on Reliability, 2009, 58(2): 295-304.
- [19] MARCELLONI F, VECCHIO M. An efficient lossless compression algorithm for tiny nodes of monitoring wireless sensor networks[J]. Computer Journal, 2009, 52(8): 969-987.
- [20] Tang Baoping, Song Tao, Li Feng, et al. Fault diagnosis for a wind turbine transmission system based on manifold learning and Shannon wavelet support vector machine[J]. Renewable Energy, 2014, 62: 1-9.

(收稿日期: 2022-04-11)

## 作者简介:

王昱钦(1987-), 男, 本科, 工程师, 主要研究方向: 数据压缩技术、无线通信传输技术、企业信息化及网络安全。

王鑫(1979-), 男, 硕士, 工程师, 主要研究方向: 无线通信技术、传感器与测控系统数据压缩技术、信息系统装备项目管理。

刘保强(1987-), 通信作者, 男, 硕士, 工程师, 主要研究方向: 通信技术、数据压缩技术、通信与信息安全等, E-mail: liubaoqiang@jari.cn。



扫码下载电子文档

## 版权声明

经作者授权，本论文版权和信息网络传播权归属于《电子技术应用》杂志，凡未经本刊书面同意任何机构、组织和个人不得擅自复印、汇编、翻译和进行信息网络传播。未经本刊书面同意，禁止一切互联网论文资源平台非法上传、收录本论文。

截至目前，本论文已经授权被中国期刊全文数据库（CNKI）、万方数据知识服务平台、中文科技期刊数据库（维普网）、DOAJ、美国《乌利希期刊指南》、JST 日本科技技术振兴机构数据库等数据库全文收录。

对于违反上述禁止行为并违法使用本论文的机构、组织和个人，本刊将采取一切必要法律行动来维护正当权益。

特此声明！

《电子技术应用》编辑部

中国电子信息产业集团有限公司第六研究所