

# 电能质量数据压缩算法研究

杨超颖

(山西电力科学研究院,山西太原 030001)

**摘要:**首先总结了常见的 4 种数据压缩基本方法,即霍夫曼(Huffman)编码、算术编码、LZ 系列编码以及基于小波分析的数据压缩法。重点分析了当前研究和应用比较广泛的 3 种数据压缩综合算法,即基于离散小波变换、多级树集合分裂(SPIHT)编码和算术编码相结合的数据压缩方法、基于 PQzip 技术的数据压缩算法和 Zlib 数据压缩算法,并比较了各自的适用条件及优缺点。最后给出了电能质量数据压缩算法选用的建议。

**关键词:**数据压缩;算术编码;SPIHT 编码;PQzip;Zlib

**Abstract:** Four common algorithms for data compression are summarized firstly, i. e., Huffman coding, arithmetic coding, LZ series coding, and data compression method based on wavelet analysis. Meanwhile, three integrated data compression methods that are current focus of the research and applied widely are analyzed, that is, data compression method based on discrete wavelet transform (DWT), set partitioning in hierarchical tree (SPIHT) coding and arithmetic coding, data compression method based on PQzip technology and data compression method with Zlib. Then, their applicable conditions, advantages and disadvantages are compared. Finally, the recommendations for selection of data compression algorithms of power quality are given.

**Key words:** data compression; arithmetic coding; SPIHT coding; PQzip; Zlib

中图分类号:TM714.3 文献标志码:A 文章编号:1003-6954(2012)01-0026-05

## 0 引言

随着电能质量监测系统规模的发展,监测点的设置越来越多,监测装置记录与分析的数据存储和有效传输已成为监测系统高效运行的瓶颈之一。为此,应采取对数据进行压缩,以减少数据存储的空间与传输量,保证传输速度,使电能质量监测系统高效运行。

数据压缩按照压缩的失真度分为有损压缩和无损压缩,有损压缩算法主要有脉冲编码、预测编码以及变换域编码算法等。考虑到数据的特殊性,在电力系统相关数据压缩中主要采用无损压缩算法,包括自适应霍夫曼编码(Huffman 编码)、LZ 系列编码、算术编码、游程编码、RAY-Period 压缩编码等算法。对于电能质量数据压缩问题的研究是近年来的一个热点。文献[1-2]基于小波变换,采取阈值方法处理小波系数,取得了一定的压缩效果。文献[3-4]基于小波包变换并结合最优小波基的选择来实现电能质量数据压缩。但上述方法得到的压缩比仍较低,维持在几倍至十几倍,且无法控制压缩性能。文献[5-6]提出了基于提升格式的电能质量数据压缩和重

构算法,提高了运算速度,适合于实时压缩。针对电力系统信号的周期性特点,文献[7-9]进一步挖掘了此类信号的循环冗余性,按整数倍周期将一维电能质量数据转换为二维矩阵,利用二维小波变换进行数据的压缩,取得了较好的压缩性能。文献[10]采用了图像压缩领域的多级树集合分裂(SPIHT)编码来实现一维数据的压缩,但没有讨论压缩码率与压缩性能的关系,且压缩性能不佳,信噪比在 30 dB 以上时,压缩比只有 30 以下。此外,PQzip 与 Zlib 算法等在电能质量数据压缩中也得到广泛应用。

将首先介绍数据压缩的 4 种基本算法,然后对目前电能质量数据压缩的综合算法进行分析,进而提出电能质量数据压缩算法选用的建议。

## 1 电能质量数据压缩基本方法

### 1.1 Huffman 编码

Huffman 编码是一种常用的压缩编码方法。它的基本原理是频繁与较少使用的数据分别用较短与较长的代码代替,每个数据的代码各不相同。

Huffman 压缩编码作为一种高效的变长编码方

式码长取决于字符的概率,在源数据中出现概率越高的符号,相应的码长越短;反之,相应的码长越长,从而在整体上缩短了源数据包含信息的比特数,达到用尽可能少的码字表示源数据。Huffman压缩编码是接近压缩比上限的一种较好的编码方法。

### 1.2 算术编码

算术编码是一种无失真的编码方法,能有效地压缩信息冗余度,属于熵编码的一种。其基本原理是:根据信息可能发现的不同符号序列的概率,把[0,1]区间划分为互不重叠可反映各符号序列概率的子区间,使每个子区间内的任意一个实数都可以用来表示对应的符号序列,这个数就是该符号序列所对应的码字。对信息进行算术编码时首先需建立信息概率表,然后需对信息发出的符号序列进行扫描编码。通过这两个过程即可实现对信息的编码压缩。

### 1.3 LZ系列编码

LZ系列算法包括LZ77、LZ78以及LZW算法等,其中LZ77和LZ78是LZ系列的基础算法,这两个算法在字典的构造和实现上代表了两种完全不同的思想<sup>[11]</sup>。

LZW的压缩基本原理是,通过分析源数据流,在编码的同时不断更新,自适应地生成一个串表,后续的数据就可以同串表中的数据相匹配。由于这个生成的串表不需要随着压缩的数据一同传输,而是能够根据压缩的数据在解压缩的时候重新动态地生成一模一样的串表,从而进一步提高了压缩的效率。

### 1.4 基于小波分析的数据压缩法

基于小波的数据压缩方法包括有多分辨率分析方法、Mallat算法、小波变换、小波包变换和多级树集合分割编码相结合的方法等。这些数据压缩方法在电力系统中也得到了广泛的应用。

其中,小波包变换能够为信号提供一种更加精细的分析方法,它将频带进行多层次划分,对多分辨率分析没有细分的高频部分进一步分解,并且能够根据被分析信号的特征,自适应地选择相应的频带,使之与信号频谱相匹配,从而提高了时-频分辨率。文献[12]详细分析了小波包变换和多级树集合分割编码相结合的方法。

## 2 电能质量数据压缩综合算法

### 2.1 基于离散小波变换、SPIHT与算术编码相结合的数据压缩方法

为满足电能质量监测网存储与传输海量录波数据的需要,文献[13]提出了提升格式的二维离散小波变换、SPIHT编码与算术编码相结合的数据压缩方法。该算法的实现原理如下图1所示。

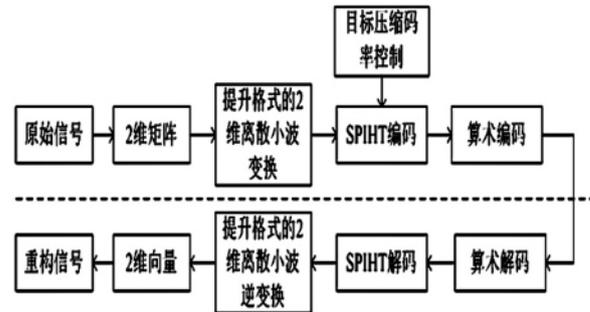


图1 电能质量数据压缩与重构方法

首先,将采集的一维电能质量数据按整数倍周期排列为二维矩阵,利用图像压缩的方法对电能质量录波数据进行压缩。在SPIHT编码中,小波系数按重要性从大到小的顺序被编码,达到目标压缩码率便立即结束编码,因此可以根据电能质量监测网的通信状况来控制压缩码率,实现对压缩性能的控制;且SPIHT编码对不重要的小波系数及其不重要的后代节点都不进行编码,提高了编码效率。最后,对SPIHT编码输出的数据进行算术编码,可进一步提高压缩比。重构是压缩的逆向过程。

对常见的电能质量扰动信号(谐波、凹陷、中断、震荡、脉冲)进行的压缩研究表明,在压缩码率 $R = 0.5$  bit/pel时,各类电能质量信号的压缩比都达到了115倍以上。在如此高的压缩比且含有10 dB强度的高斯白噪声的情况下,重构信号的信噪比也达到了32 dB以上。

另外,此数据压缩方法能灵活地调节压缩性。压缩码率越高,压缩比越低,信噪比越高,均方误差也越小。因此,可以根据电能质量监测网的通信状况,通过控制压缩码率来调节压缩比。当通信状况较好时,可提高压缩码率,传输较多的数据量,此时重构信号的信噪比较高,能较准确地记录原始信号的特征;当通信状况较差时,可降低压缩码率,传输较少的数据量,此时重构信号的信噪比较低,但仍能记录原始信号的大概特征。

### 2.2 基于PQzip技术的数据压缩算法

PQzip是一种先进的电能质量数据压缩技术,主要应用于以色列Elspec公司开发的G4400电能质量监测装置中。PQzip技术中所采用的基于FFT变换

的有损压缩算法是当今应用广泛的一种方法,该方法精度较高,功能较多,使用方便,特别适用于正弦波的情况。但FFT是在整个时域内的积分,只能适用于确定性的平稳信号(如谐波),对时变非平稳信号却难以充分描述。

基于PQzip技术,文献[14]采用FFT和小波包变换两种方法相结合实现了数据压缩。对于染噪信号,为有效提取信号的特征量,首先利用小波阈值法去噪,再对其进行二进小波变换检测信号:如果检测到的信号为稳态或稳态扰动,采用FFT变换对信号进行压缩,得到相应的频谱,记录原始信号中不同的频率分量;如果检测到的信号为暂态扰动,则采用小波包变换进行压缩,对小波包各层分解系数进行阈值处理,保留与信号奇异性相关的点,忽略一部分与信号奇异性无关的点。最后利用LZW编码对经过有损压缩后存储的数据进一步无损压缩,实现更有效的压缩效果。

对5种常见的暂态扰动信号以及4种稳态扰动信号的仿真分析表明<sup>[14]</sup>,对于同一稳态扰动信号,PQzip的压缩率(CR)最小,压缩效果最好,而小波变换算法CR明显大于PQzip和所提算法,说明了FFT算法在稳态以及稳态扰动信号压缩方面具有无可比拟的优越性;对于同一暂态扰动信号,小波算法的CR最大,压缩效果最差,PQzip所采用的FFT算法和所提算法的CR较为接近,这是由于小波包变换对扰动数据进行压缩时,不仅对小波包低频系数进行阈值处理,而且对高频系数同样进行阈值处理,是一种比小波变换更为精细的变换方法,其压缩效果优于小波变换的压缩方法。

电能质量信号在不同方法下的均方根误差(RMSE)结果分析表明,PQzip所采用的FFT算法在稳态信号压缩时,RMSE很小,保持在 $10^{-4}$ 数量级以上,但在处理暂态信号时,压缩率只有 $10^{-3}$ 数量级;小波变换算法在进行压缩时,RMSE基本都保持在 $10^{-3}$ 数量级;而所提算法的RMSE在处理无论是稳态还是暂态信号时,RMSE都保持在 $10^{-4}$ 数量级上,压缩后的重构信号可以较好地保留原始信号的特性。综上分析可得,电能质量信号不管是稳态、稳态扰动还是暂态扰动,采用所提算法,即基于FFT和小波包变换压缩法,在保证原始信号局部特性的基础上,能有效地压缩数据,并且取得令人满意的压缩效果。

### 2.3 Zlib数据压缩算法

Zlib压缩算法是一种基于字典的无损压缩算法,应用于PQDIF压缩数据。该算法采用了统计模型和字典模型,有机地结合了Huffman树编码法和LZ77编码法的优点,使压缩效果更好。

Zlib算法继承了字典压缩算法的思想,将此思想与滑动窗口(在内存中开辟的一个固定长度的缓冲区)相结合,把滑动窗口看成是字符的字典,找出超前察看缓冲区(与滑动窗口的缓冲区相邻)中能与滑动窗口的字符串相匹配的最长的字符串,并将此串按照LZ77编码法进行编码。当两个缓冲区中没有相匹配的字符时,便采用自适应的Huffman编码法进行编码,在编码的过程中将已编好的压缩码实时地写入压缩文件中。也就是说,Zlib算法将长度不同的符号串编码成一个个新单词,形成一本短语词典的索引,然后用索引来检索数据,并用字典的编码代替符号串以达到压缩目的。解码过程是编码的逆过程,同样用到了Huffman算法和LZ77算法的编码算法,解码过程与编码过程相似。

根据某省级数据中心建设规划,为了最大限度地保护现有投资,该省将PQView数据库作为基础数据库,PQView软件本身提供了数据采集功能、数据管理功能及数据分析功能。PQView数据库中数据存储采用PQDIF格式,对稳态数据及事件的录波数据进行压缩的算法采用ZLib,此算法的压缩效率非常高。对于纯数据文件2.04 MB,压缩后13.5 kb,压缩率为99.4%;纯数据文件1.00 MB,压缩后7.23 kb,压缩率为99.3%;纯数据文件454 kb,压缩后3.81 kb,压缩率为99.2%;纯数据文件221 kb,压缩后为99%;对于文本+数据文件,原文件大小为228 kb,压缩后2.62 kb,压缩率为99%;原文件大小为3.14 MB,压缩后为21.4 kb,压缩率为99.4%。文件的容量越大,ZLib算法的效率就越高。

### 2.4 电能质量数据压缩其他综合算法

目前,对电能质量数据压缩方法的研究有很多,文献[15]基于二维DCT的电能质量监测数据压缩方法,该方法仿真结果表明,当均方误差为3.89%时,压缩比可以达到82.8%。文献[16]基于混合小波包的电能质量数据压缩算法,对于电能质量事件数据的压缩性能优于传统小波/小波包变换。文献[10]基于整数小波变换和SPIHT编码的录波数据压缩算法,其整数小波变换运算速度快,节约内存,易于DSP实现;SPIHT编码方法形成的码流易于实现可变码

率,也能达到一定的压缩性能。文献 [17] 基于 PQDIF 格式压缩的电能质量数据模型,是基于 Zlib 的数据压缩技术,采用此数据模型基于 Web 技术的电能质量监测和分析系统已投入运行。这些数据压缩法都具有自己的优缺点和使用条件。第三节着重分析比较目前被投入实际使用和研究广泛的 3 种数据压缩方法。

### 3 电能质量数据压缩 3 种方法的分析比较

基于二维离散小波变换、SPIHT 并结合算术编码方法的数据压缩方法对电能质量数据进行压缩。在较高的压缩比时仍能保证较高的信噪比,信号的特征基本得到了保留;压缩性能可以根据电能质量监测网的通信状况灵活调节,避免了因为通信线路上传输的数据量过大而造成的数据丢失或严重滞后,监测中心能及时地监测、分析电能质量状况。提出的方法能较好地解决电能质量监测网对海量录波数据进行存储与传输的问题。

基于 PQzip 技术的数据压缩算法结合了 FFT 和小波包变换压缩的优点,即 FFT 具有良好的频域特性和小波变换具有更高的时频分辨率,并有很好的抗噪性,可以自动检测扰动信号并进行分类:对稳态以及稳态扰动信号采用 FFT,对暂态信号采用小波包变换进行压缩。与 PQzip 以及小波变换算法相比较,抗噪性能好,还可以在保证原始信号局部特性的基础上,获得更好的压缩效果,CR 较小,并且其 RMSE 也保持在 10<sup>-4</sup> 级别以上,取较满意的压缩结果。该算法在实际应用中,若忽略扰动、畸变和直流分量的影响,在工程运用中,可以广泛采用。

基于 Zlib 的数据压缩法是一项成熟的压缩算法,Zlib 是一个免费、通用、不受任何法律阻碍的、无损的数据压缩开发库,而且还是跨平台的。与其他无损压缩算法相比,Zlib 压缩算法能够压缩更多形式的数据,并且具有同 winzip 和 winrar 等商业软件相比毫不逊色的压缩率,并且使用了很少的系统资源,对各种数据提供很好的压缩效果。如果经常使用 Zlib 库压缩特定类型的数据,可以使用有针对性的策略,提高压缩效率和性能。尽管 Zlib 只支持 deflate 压缩算法,压缩倍数比所提到的其他算法大,但在实际运用中还是很广泛,运用方便简单。

## 4 结 语

在简要总结几种典型的电能质量数据压缩基本方法的基础上,着重分析了目前研究和较为广泛的数据压缩法,并比较了各个方法的适用条件及优缺点。若以较少的数据质量损失存储大量的数据,从而获得很高的压缩率,若考虑使用有损压缩,忽略压缩过程中损耗的数据,建议考虑使用 PQzip 数据压缩法,该方法能提供过去的所有与电能质量相关的数据,不论其重要程度怎样,任何时间的数据都被存储;若考虑使用无损压缩算法,根据以上的分析,建议使用 Zlib 库,对数据进行压缩,这是目前使用比较广泛成熟的压缩方法;若考虑压缩性能中的压缩比,以及误差精度,所提的基于离散小波变换、SPIHT 编码、算术编码相结合数据压缩法是目前数据压缩研究领域取得较好效果的方法。

这 3 种数据压缩法都有各自的特点,根据不同的实际需求,选择适合的电能质量数据压缩法,达到预期的数据压缩效果。但它们都有各自的影响因素,如 PQzip 数据压缩法,实际电力系统中,电能质量信号会存在各种扰动和畸变,有时还会受到直流分量的影响,而且这些扰动和畸变很多时候会同时出现。在实际应用时,影响数据压缩的因素有数据量、采样率、电能质量问题和 AD 性能等,都应当在算法中将这些因素都考虑进去,使得算法更加合理实用<sup>[18]</sup>。因此在对各种数据压缩算法进行改进研究时要综合考虑各方面的影响因素从而达到提高压缩仪器精确度的目的。

### 参考文献

- [1] P. K. Dash, B. K. Panigrahi, D. K. Sahoo, et al. Power Quality Disturbance Data Compression, Detection, and Classification Using Integrated Spline Wavelet and S-Transform [J]. IEEE Trans. on Power Delivery, 2003, 18(2): 595-600.
- [2] 刘应梅,白晓民,易裕,等.基于最小极大法的电力扰动信号压缩[J].电网技术,2004,28(3):33-37.
- [3] 成敬周,张举勇.基于小波包分析的录波数据压缩原理[J].电力自动化设备,2004,24(7):53-56.
- [4] 潘文霞,李春林,史林军.基于最佳小波包基的电能质量暂态数据压缩[J].电力系统及其自动化学报,2005,17(3):50-54.
- [5] 闫常友,杨奇逊,刘万顺.基于提升格式的实时数据压缩和重构算法[J].中国电机工程学报,2005,25(9):6-10.
- [6] 王学伟,孙金凤,王琳.基于整数小波变换的 Ray-Period 压缩算法[J].电网技术,2008,32(7):46-50.

- [7] mer Nezhir Gerek ,Dogan Gökhan Ece. 2 - D Analysis and Compression of Power Quality Event Data [J]. IEEE Trans. on Power Delivery 2004 ,19( 2) : 791 - 798.
- [8] mer Nezhir Gerek ,Dogan Gökhan Ece. Compression of Power Quality Event Data Using 2D Representation [J]. Electric Power Systems Research 2008 ,78( 6) : 1047 - 1052.
- [9] 鲍文,周瑞,刘金福. 基于二维提升小波的火电厂周期性数据压缩算法[J]. 中国电机工程学报,2007,27(29):96-101.
- [10] 王永玉,孙衢,袁超伟. 基于整数小波变换和SPIHT编码的录波数据压缩算法[J]. 继电器 2008,36(11):50-54.
- [11] 郝蔓钊. 电能质量监测数据的数据压缩[D]. 上海:上海交通大学,2009.
- [12] 高山. 基于小波包变换的多级树集合分裂编码算法研究[D]. 西安:西安科技大学,2009.
- [13] 刘毅,周维维,冯浩. 电能质量监测网录波数据压缩方法[J]. 电网技术,2010,34(4):199-203.
- [14] 柯慧. 电能质量数据压缩算法研究[D]. 上海:上海交通大学,2009.
- [15] 胡志坤,何志敏,安庆,等. 基于二维DCT的电能质量监测数据压缩方法[J]. 中南大学学报,2011,42(4):1021-1027.
- [16] 郑伟彦,吴为麟. 基于混合小波包的电能质量数据压缩算法[J]. 电工电能新技术,2010,29(3):8-11.
- [17] 丁屹峰,程浩忠,占勇,等. 基于PQDIF格式压缩的电能质量数据模型[J]. 继电器,2005,33(7):55-58.
- [18] 朱宏涛. 电网数据采集系统建模与数据压缩品质影响的研究[D]. 哈尔滨:哈尔滨工业大学,2010.

(收稿日期:2011-10-08)

(上接第25页)

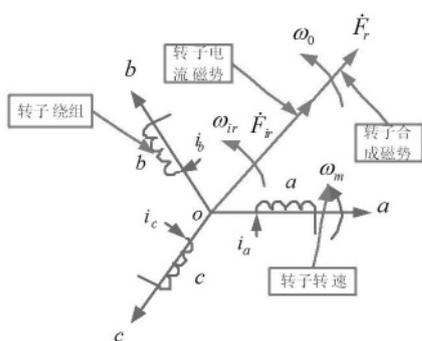


图4 正常时转子合成磁势的原理

发电机的并网故障如图5所示,转子合成磁势

$\dot{F}_r$  的频率约为  $f = 43.33 - 6.67 = 36.66$  Hz,感应出的定子电压频率约为 36.66 Hz。

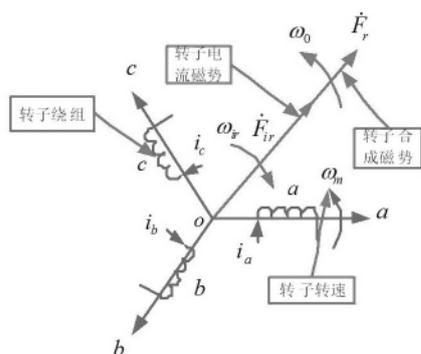


图5 非正常时转子合成磁势的原理

此时,只要调换任意两根转子接线,就可使转子

合成磁势  $\dot{F}_r$  频率正常。

## 2.5 相位不同步的原因

定子电压的相位控制是分两步完成的。

第一步,通过PLL测量电网电压的相角  $\theta_u$ ,计算转子旋转后的相角  $\theta_2'' = \int_0^t \omega_m dt$ ,使得转子电压相角

$\theta_{ru} = \theta_u + \pi/2 - \theta_2''$ ,这就是频率控制的原理。

第二步,计算出定子电压的相角  $\theta_{su}$ 、电网电压的相角  $\theta_u$ ,然后取其差值  $\theta_{offset} = \theta_u - \theta_{su}$  即是补偿相角。取  $\theta_{ru} = \theta_u + \pi/2 - \theta_2'' + \theta_{offset}$ ,这个相角即可使定子电压的相位同步。

变流器从第一步进入第二步,是有条件的,那就是变流器先检测定子电压的  $d$ 、 $q$  分量是否满足要求。

从轴伸端看,B厂家的发电机转子绕组  $a-b-c$  为顺时针排列,定子电压为负序,经  $d-q$  变换后的为二倍频正弦,不能满足变流器进入第二步的条件。变流器始终停留在第一步,而不能进入第二步进行相位同步操作。任意调换定子的两根接线后,相位不同步的问题得以解决。

## 3 结 语

分析了一例双馈发电机的实验故障,总结出变流器的软件配置与双馈发电机配置两者的兼容非常重要。如果变流器的软件配置与发电机的配置不兼容,可能导致发电机无法并网,或者并网后对发电机造成损害。

## 参考文献

- [1] 苑国锋,柴建云,李永东. 变速恒频风力发电机组励磁变频器的研究[J]. 中国电机工程学报,2005,25(8):90-94.
- [2] 郎永强. 交流励磁双馈电机风力发电系统控制技术研究[D]. 哈尔滨:哈尔滨工业大学,2007.
- [3] 王永,沈颂华,关森. 新颖的基于电压空间矢量三相双向整流器的研究[J]. 电工技术学报,2006,21(1):104-110.

(收稿日期:2011-08-31)