文章编号:1671-251X(2013)10-0107-04　　DOI:
张卫国. 数据压缩算法在煤矿类软件系统中的应用研究[J]. 工矿自动化,2013,39(10):107-110.

数据压缩算法在煤矿类软件系统中的应用研究

张卫国

(天地(常州)自动化股份有限公司,江苏 常州　213015)

摘要:为了解决煤矿类软件系统在大规模数据量情况下的数据高效存储和查询问题,提出了一种将旋转门 SDT 数据压缩算法应用于煤矿类软件系统的方案。介绍了该算法的原理及应用推导,给出了模拟量和开关量压缩数据解压还原方法和特殊情况处理方法。应用结果表明,该算法压缩比高,节省了大量的存储空间,能准确地反映出数据的走势和变化。

关键词:过程数据;数据压缩;SDT 算法;解压还原
中图分类号:TD672　　文献标志码:A　　网络出版时间:
网络出版地址:

Application research of data compression algorithm in coal mine software system

ZHANG Wei-Guo
(Tiandi (Changzhou) Automation Co., Ltd., Changzhou 213015, China)

Abstract: In order to solve problem of efficient data storage and query under condition of large amount of data in coal mine software system, a scheme of applying SDT data compression algorithm into coal mine software system was proposed. Principle and application derivation of the algorithm were introduced, and decompression method and treatment method under special circumstances of analog and switching compression data were given. The application result shows that the algorithm has high compression ratio, saves great storage space, and can reflect trend and change of data.

Key words: process data; data compression; SDT algorithm; decompression

0　引言

面对煤矿生产规模的逐步扩大,在数据自动化采集过程中涉及到的数据也在成倍增长。面对采集到的大规模监控类数据量,如何合理、高效地进行数据存储及查询是当前煤矿类软件系统需要处理的一个关键问题,必须考虑对数据进行压缩处理。

由于生产过程实时数据都是秒级数据,如果采用传统的直接存储方案,将会占用大量的存储空间,大大增加存储成本[1]。生产过程采集到的原始数据中包含大量的冗余和不相关信息,在保留关键特征数据的基础上去除冗余和不相关信息,就可达到数据压缩的目的。数据压缩的本质是数据变换,它试图在压缩比和信号保真度之间寻求一种折衷。过程数据压缩方法基本分为 3 种[2-3],即分段线性方法、矢量量化方法及信号变换法[4]。而分段线性方法又包括矩形波串法、向后斜率法、SDT(Swing Door Trending,旋转门)法[5]及 PLOT(Piecewise Linear On-line Trending,分段线性在线趋势)法。

煤矿生产过程中采集到的数据大多是缓慢变化的数据,采用分段线性化的方法可以很好地表示历史数据的变化趋势。而在实际应用中,SDT 数据压缩算法是分段线性方法中比较常用的一种方法,其突出优点是算法简单、执行速度快。本文对该数据压缩算法的原理及应用推导步骤进行分析,给出了压缩数据解压还原的方法,并结合实际应用项目数

收稿日期:2013-07-19。
基金项目:科技部院所技术开发专项基金项目(2012EG122196);天地(常州)自动化股份有限公司研发项目(12SY005)。
作者简介:张卫国(1982—),男,河南洛阳人,工程师,硕士,主要从事煤矿系统相关软件的设计与研发工作,E-mail:28120079@qq.com。

据给出压缩效果，证明了该算法的有效性。

1 SDT 数据压缩算法原理

SDT 数据压缩算法基本原理如图 1 所示。数据点 a 上下各有一点，它们与数据点 a 之间的距离为 E，这 2 点作为“门”的 2 个支点。当只有第 1 个数据点时，2 扇“门”是关闭的。随着采集数据点的增加，“门”的敞开角度逐步增大。每扇“门”只能向更大的敞开角度打开而不能闭合，只要 2 扇“门”未达到平行，即2 个内角之和小于 180°，这种“转门”操作就可以继续进行。图 1 中的数据点 a,b,c,d,e 经过 SDT 数据压缩算法处理后，将只存储数据点 a 和数据点 e；从点 e 开始，2 扇“门”再恢复到初始状态，开始时 2 扇“门”关闭，然后逐步打开，到数据点 h 后 2 扇“门”再次平行，所以数据点 e,f,g,h 经过 SDT 数据压缩算法处理后，将只存储数据点 e 和数据点 h，当然数据点 e 在前一处理步骤中已经存储，将不再重复存储。

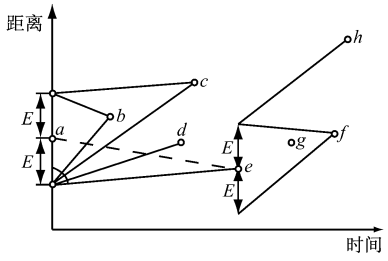

图 1 SDT 数据压缩算法基本原理

2 SDT 数据压缩算法的应用推导

SDT 数据压缩算法的推导步骤如图 2 所示。2 扇“门”fa 和 ga 在经过数据点 b,c,d 的 SDT 数据压缩算法处理后，两者所处的开合状态为 fd 和 gc。对于数据点 d，是否满足“转门”操作的要求，判断标准为 $\alpha+\beta<180°$。

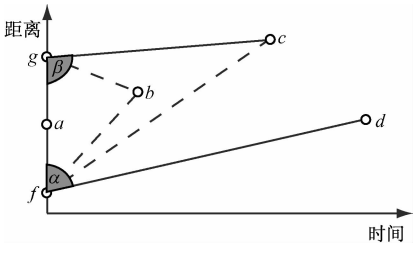

图 2 SDT 算法步骤

设点 f 坐标为 $(0,f_y)$，点 g 的坐标为 $(0,f_y+2E)$，点 c 的坐标为 (c_x,c_y)，点 d 的坐标为 (d_x,d_y)。则有

$$\alpha=90°-\arctan[(d_y-f_y)/(d_x-0)]$$

$$\beta=90°+\arctan[(c_y-f_y-2E)/(c_x-0)]$$

$$\alpha+\beta<180°$$

$$\Leftrightarrow 90°-\arctan[(d_y-f_y)/(d_x-0)]+90°+\arctan[(c_y-f_y-2E)/(c_x-0)]<180°$$

$$\Leftrightarrow \arctan[(c_y-f_y-2E)/(c_x-0)]<\arctan[(d_y-f_y)/(d_x-0)]$$

$$\Leftrightarrow (c_y-f_y-2E)/(c_x-0)<(d_y-f_y)/(d_x-0)$$

根据以上推导结论，即可通过编程实现 SDT 数据压缩算法。

模拟量数据可采用 SDT 算法进行数据压缩；开关量数据在数据值发生变化时进行存盘；对于累计量，可在本次值与上次值之差小于零时存储上次累计量的值。

在实现 SDT 数据压缩算法进行数据压缩的过程中，模拟量数据压缩还要在 SDT 算法的基础上增加对特殊情况的处理：

(1) 模拟量数据状态为故障等的处理。无论模拟量数据在发生故障前的一次采样数据是否满足 SDT 算法进行存储的条件，都需要对该采样数据进行存储。对于故障的第一个采样数据也要作为压缩数据进行存储，在故障结束后的第一个采样数据也要作为压缩数据进行存储。

(2) 模拟量采样值突变情况的处理。当模拟量值发生突变时，为了使压缩后的数据能正确反映出数据突变的情况，需要将突变的采样值、突变前的一个采样值作为压缩数据进行存储。

3 压缩数据的解压还原

当需要查看指定测点的历史数据时，就需要对相关存储数据进行解压还原。数据的解压还原可以通过采取直线趋势化逼近的方法进行还原，该过程相对简单，但模拟量测点和开关量测点要分开进行还原处理。

3.1 模拟量测点数据还原方法

需要还原的时间段为 StartTime 到 EndTime，首先到数据库中查找出该测点在 StartTime 到 EndTime 内所有数据的集合 DataSet{data(1),data(2),…,data(N),…,data(end)}，各个数据对应的时间集合 TimeSet 为{time(1),time(2),…,time(N),…,time(end)}。如果需要还原出时间点 time(X) 对应的数据 RestoreData(X)，假设 time(X)>time(N)，time(X)<time($N+1$)，则 RestoreData(X)=[time(X)−time(N)]×[data($N+1$)−data(N)]/[time($N+1$)−time(N)]+data(N)。

如果需要还原的时间点 time(X)>StartTime,且 time(X)<time(1),则对应的还原数据 RestoreData(X)有 2 种处理方法:

(1) 可以简单的处理,即 RestoreData(X)=data(1)。

(2) 从数据库中查出该测点在时间 time(1)之前的第 1 条数据 data(0),其对应时间为 time(0),则 RestoreData(X)=[time(X)−time(0)]×[data(1)−data(0)]/[time(1)−time(0)]+data(0)。

对于 time(X)>time(end),且 time(X)<EndTime 的情况也需做类似处理。

特殊情况处理:

如果从数据库中查找出该测点在 StartTime 到 EndTime 内所有数据的集合 DataSet 为空,则需要从数据库中查找出该测点在时间 StartTime 之前的第 1 条记录 Data(start),其对应时间为 Time(start),以及该测点在时间 EndTime 之后的第 1 条记录 Data(end),其对应时间为 Time(end)。最后,对于需要还原的时间点 Time(X),其对应的还原数据为 RestoreData(X)=[time(X)−time(start)]×[data(end)−data(start)]/[time(end)−time(start)]+data(start)。

3.2　开关量测点数据还原方法

需要还原的时间段为 StartTime 到 EndTime,首先到数据库中查找出该测点在 StartTime 到 EndTime 内所有数据的集合 DataSet{data(1),data(2),…,data(N),…,data(end)},各个数据对应的时间集合 TimeSet 为{time(1),time(2),…,time(N),…,time(end)}。如果需要还原出时间点 time(X)对应的数据 RestoreData(X),假设 time(X)>time(N),time(X)<time(N+1),则 RestoreData(X)=data(N)。

如果需要还原的时间点 time(X)>StartTime,且 time(X)<time(1),则对应的还原数据 RestoreData(X)的处理方法:从数据库中查出该测点在时间 time(1)之前的第 1 条数据 data(0),其对应时间为 time(0),则 RestoreData(X)=data(0)。对于 time(X)>time(end),且 time(X)<EndTime 的情况,则 RestoreData(X)= data(end)。

特殊情况处理:

如果从数据库中查找出该测点在 StartTime 到 EndTime 内所有数据的集合 DataSet 为空,则需要从数据库中查找出该测点在时间 StartTime 之前的第 1 条记录 Data(start),其对应时间为 Time(start),以及该测点在时间 EndTime 之后的第 1 条记录 Data(end),其对应时间为 Time(end)。如果 2 条记录均能查到,则对于需要还原的时间点 Time(X),其对应的还原数据为 RestoreData(X)=Data(start)。

4　数据压缩效果实验

笔者编写相关实验程序对煤矿现场的瓦斯采集数据进行了压缩分析实验。现场某瓦斯传感器的原始历史数据曲线如图 3 所示,其中该瓦斯传感器的原始数据条数是 19 212 条。

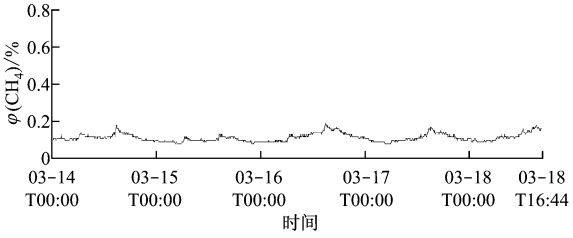

图 3　原始历史数据曲线

该瓦斯传感器的原始历史数据按 SDT 压缩算法进行处理后的压缩数据曲线($E=0.01$)如图 4 所示,其中该瓦斯传感器原始历史数据压缩处理后的数据条数是 777 条。

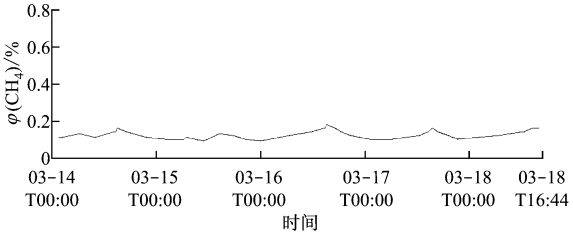

图 4　压缩数据曲线

从图 3、图 4 可看出,该瓦斯传感器的原始采样数据曲线和压缩处理后的数据曲线整体走势一致,压缩后的数据曲线能够较为正确地反映出该瓦斯传感器的历史数据信息;该瓦斯传感器的原始数据条数是 19 212,当 $E=0.01$ 时压缩后的数据条数为 777,压缩比率为 4.04%,实现了对原始数据较高的压缩比,节省了大量的存储空间。

5　结语

通过对 SDT 数据压缩算法的应用推导,并结合煤矿类软件系统中采样数据的具体特点,给出了 SDT 数据压缩算法在煤矿类软件系统中的具体应

用实现。实验结果表明,压缩后的数据曲线能较为准确地反映出数据的走势和变化,并有效减少数据存储量。将 SDT 数据压缩算法在现有的大数据量煤矿类数据采集软件系统中加以应用,将能有效解决大数据量存储占用空间资源大、耗时长的问题,从而在降低软件系统成本的同时提高用户的使用满意度。

参考文献:

[1] 段培永,姚庆梅,汤同奎. 一种用于过程数据压缩的矩形波串法及其性能分析[J]. 厦门大学学报:自然科学版,2001,40(增刊 1):265-268.

[2] 段培永,汤同奎,邵惠鹤. 一种新的 Boxcar 算法及其在数据压缩中的应用[J]. 计算机应用,2001,21(9):14-17.

[3] 董栋,刘强. 一种改进的分段线性趋势压缩算法及应用[J]. 微计算机信息,2006,22(36):200-202.

[4] 赵利强,于涛,王建林. 基于 SQL 数据库的过程数据压缩方法[J]. 计算机工程,2008,34(14):58-59.

[5] 齐文斌,李东平,杨东,等. 广域测量系统数据在线无损压缩算法[J]. 电网技术,2008,32(8):86-90.