

# 城市轨道交通列车运行数据实时压缩方案

李德祥 王林美 王 曙

(中车青岛四方车辆研究所有限公司, 266114, 青岛)

**摘 要** [目的] 针对目前城市轨道交通列车运行数据存在的磁盘存储空间需求增加、磁盘频繁损伤及高带宽占用等问题, 研究数据压缩对于减少存储空间需求, 缓解磁盘负担, 以及减少带宽浪费具有重要意义。[方法] 分析了列车运行数据的特性, 从数据自身的时间相关性、空间相关性及整体冗余性等三个维度对其内部的冗余特性进行研究, 提出帧间压缩 + 改进的 RLE(游程编码) 压缩 + ZIP 压缩的多维度实时压缩方案以降低数据的冗余性。通过算法的实时性验证试验, 以及实际列车运行数据验证试验, 证明此压缩方案的有效性。[结果及结论] 提出的数据多维度实时压缩方案复杂度低, 且具备压缩能力强、压缩速度快的优点。试验及应用表明, 所提方案的数据实时压缩操作耗时仅为 0.16 ms, 满足 20.00 ms 的数据采集周期; 所提方案的最终压缩率可达 3.35%, 同样成本的硬件能够存储更长时间数据; 最终写入存储磁盘和无线传输的数据量缩减至原数据的 11.83%, 大幅减少了磁盘擦写的次数和无线传输的数据量, 有效降低了磁盘存储故障率以及无线传输带宽占用。

**关键词** 城市轨道交通; 列车运行数据; 数据实时压缩

**中图分类号** TP274

DOI:10.16037/j.1007-869x.2025.01.046

## Real-Time Compression Scheme for Urban Rail Transit Train Operation Data

LI Dexiang, WANG Linmei, WANG Shu

(CRRC Qingdao Sifang Rolling Stock Research Institute Co., Ltd., 266114, Qingdao, China)

**Abstract** [Objective] In view of the current challenges such as increased requirements for disk storage space, frequent disk damage and high bandwidth occupation existing in urban rail transit train operation data, research on data compression is of great significance for reducing storage space requirements, alleviating disk burden, and cutting bandwidth waste. [Method] By analyzing the characteristics of train operation data, the internal redundant characteristics of the data are studied from three dimensions i. e. time correlation, spatial correlation and overall redundancy of the data itself. A multi-dimensional real-time compression scheme with inter-frame compression, improved RLE (Run-Length Encoding) compression and ZIP compression is proposed to reduce the data redundancy. The

scheme is proved to be effective by the real-time verification test of algorithms and the verification test of the actual train operation data. [Result & Conclusion] The proposed multi-dimensional real-time data compression scheme features low complexity, strong compression ability and fast compression speed. Tests and applications show that it takes only 0.16 ms for the real-time data compression with the proposed scheme, meeting the 20.00 ms data acquisition cycle requirement. The final compression ratio with the proposed scheme can reach 3.35%, and hardware of the same cost can store data for a longer time. The final amount of data written to storage disks and wirelessly transmitted is reduced to 11.83% of the original one, greatly reducing the times of disk erasures and the amount of data wirelessly transmitted, and effectively lowering the disk storage failure rate and the occupancy of wireless transmission bandwidth.

**Key words** urban rail transit; train operation data; real-time data compression

列车运行数据(简称“行车数据”)指列车运行中各类状态、故障等信号,是智能运维、故障分析的基础。目前,针对行车数据在记录和传输过程中的压缩技术研究较少。文献[1]提出引入数据变化标记位,当数据变化时再进行记录,但随着记录数据增加,标记位也占用较多空间。有大量应用对数据采用不压缩实时传输<sup>[2]</sup>;也有应用为了减少带宽占用,待列车回库再对数据进行批量压缩传输<sup>[3]</sup>,此方法不能实时地减少数据记录和传输的数据量。文献[4]直接采用 LZO(Lempel-Ziv-Oberhumer)压缩算法对数据进行压缩,以节省存储空间,提升传输效率。本文对目前行车数据记录存在的问题进行了阐述,基于行车数据特性分析对数据压缩方法进行研究,并通过试验和应用验证了方案的有效性。

## 1 行车数据特征分析

用于记录和传输的行车数据从早期几百 MiB 快速增长到几十 GiB,随之而来的问题有:存储 falsh

因成本限制无法持续增长,存储 flash 长期频繁擦写导致记录设备出现系统故障<sup>[5]</sup>,车地数据大量传输导致高带宽占用、高通信成本等问题。

上述问题可通过降低数据存储量和传输量有效解决上述问题,同时需满足如下条件:

- 1) 压缩能力强。减少 flash 擦写次数,降低通信成本和带宽占用。
- 2) 速度快。保证数据传输和记录的实时性。
- 3) 准确还原数据。须采用无损压缩算法,若压缩数据一旦出现问题,应尽可能地缩小影响范围。

设备按固定时间间隔采集 1 帧数据进行记录和传输,数据帧格式为:[帧头,时间,信号 1,信号 2,⋯,信号  $n$ ,校验,帧尾]。将一段时间数据看作二维矩阵,则 1 行表示同一时刻不同信号值的空间排列,1 列为同一信号不同时刻值的排列。

从数据的空间维度分析,1 行(1 帧)数据存在如下冗余:

- 1) 物理冗余。同物理含义信号多处采集。例如,列车速度,电机速度及车轴速度虽采集位置不同,但数值基本一致。
- 2) 数字量冗余。数据中的数字信号以 0 和 1 两种状态表示,且存在大量的状态重复,即信号冗余。
- 3) 状态冗余。信号反映列车状态,例如:列车静止时,牵引系统相关信号值大部分为 0;降弓后,辅助逆变相关信号值大部分为 0。

从数据的时间维度分析,1 列数据为同一信号的时间序列。列车运行时,信号变化平滑;列车静止时,信号可能长时间不变,相邻数据帧存在大量重复值。

数据记录形成文件后,内部仍有较多冗余的字符,例如帧头、时间及帧尾等。因此,从文件整体角度,也可以继续降低数据冗余。

## 2 行车数据压缩研究

本文基于行车数据特征分析,提出一种数据多维度实时压缩方案,分别从时间排列、空间排列及整体等 3 个角度研究了压缩方法,并提供了参数取值建议。

### 2.1 降低数据时域冗余

考虑数据帧之间时域上的强相关性,文献[6]参考 MPEG(动态图像专家组)编码提出帧间压缩方法并应用于遥测数据压缩。该压缩方法较为复杂,本文给出更为简便的方式。该压缩方法步骤如下:

步骤 1 引入参数  $K$ ,每隔  $m$  帧取 1 帧数据作为关键帧 I。

步骤 2 将两个 I 帧间的所有帧定义为非关键帧 P,与 I 帧对比,将不一致的数据记为〈不一致字节数,起始偏移,不一致数据段〉。为了说明该压缩方法,表 1 列举了采集的行车数据示例。

表 1 采集的行车数据示例

Tab.1 Examples of collected train running data

帧序 号	不同字节下的行车数据																		
	0	1 B	2 B	3 B	4 B	5 B	6B	7 B	8 B	9 B	10 B	11 B	12 B	13 B	14 B	15 B	16 B	17 B	18 B
0	a	a	b	c	c	c	c	c	c	i	k	k	k	k	n	o	p	d	e
1	a	a	b	c	e	f	g	h	i	i	k	k	d	k	n	o	p	d	e
2	a	t	m	k	c	c	c	c	c	i	k	e	k	k	n	o	p	d	f

注:表中 1、2 帧与 0 帧不一致的数据加粗显示;a—i, k, m—p 代表不同的数据。

选择表 1 中第 0 帧数据为关键帧,则第 1 帧数据压缩后标记为:〈5,4,efghi〉,〈1,12,d〉,第 2 帧数据压缩后标记为:〈3,1,tmk〉,〈1,11,e〉,〈1,18,f〉。

步骤 3 考虑到 1 帧数据可能有几千字节,引入 3 个标记字节,高 11 位用于存放“不一致长度”,范围为 0~2 048,低 13 位用于存放“起始偏移”,范围为 0~8 192。将步骤 2 中的标记进行编码:第 1 帧数据压缩为 0,0xA0,4,e,f,g,h,i,0,0x20,12,d;

第 2 帧数据压缩为 0,0x60,1,t,m,k,0,0x20,11,e,0,0x20,18,f。

针对关键帧间隔参数  $K$ ,由于行车数据变化频率较缓,当实测  $1 < K < 100$ ,随着  $K$  增大,压缩率(压缩后数据大小/原数据大小×100%)持续降低,代表压缩能力持续提升。 $K$  和压缩率关系曲线如图 1 所示。

根据图 1,仅考虑提高压缩能力, $K$  的取值应增

大,但由于非关键帧的解压完全依赖于关键帧 I,一旦某关键帧在传输或记录过程中出现问题,所有对应的非关键帧将无法恢复。即:随着  $K$  增大,在压缩率提升的同时,数据损失风险也在增加。

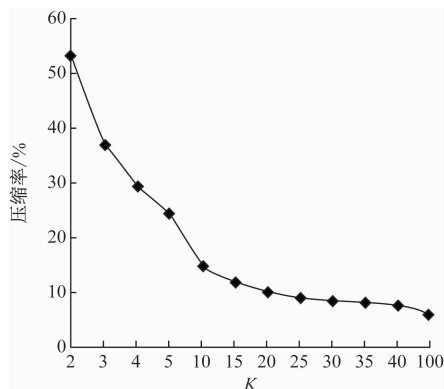


图 1  $K$  和压缩率关系曲线

Fig. 1 Relationship curve between parameter  $K$  and compression ratio

基于工程经验,本文建议取  $K = 1$  s/采集周期,即最多允许损失的数据段时长为 1 s。

## 2.2 降低数据空间方向冗余

本文对非关键帧数据进行了压缩,为进一步降低数据量,应对关键帧数据也进行压缩。I 帧含有大量重复数据,可以采用无损压缩算法。典型的算法包括 Huffman 编码、RLE(游程编码)及基于字典的 LZ(Lempel-Ziv)压缩的系列算法。Huffman 编码在压缩、解压时需要构造 Huffman 树,内存占用高且速度慢。LZW(Lempel-Ziv-Welch)压缩算法是 LZ 压缩算法系列中最常用的算法<sup>[7]</sup>,压缩、解压时能够动态构造字典且无须存储。RLE 算法简单,压缩、解压速度快,对连续重复的数据有很高的压缩比<sup>[8]</sup>。本文选用 LZW 压缩算法和 RLE 压缩算法进行对比分析。

本文对 RLE 压缩算法进行了简单改进,将多个字节的压缩标记融合成 1 B 压缩标记,最高位标记是否压缩,低 7 位表示字符重复次数或未压缩字符数量,范围为 0~127。

为对比两种算法效率,使用某城市轨道交通 1 列列车约 0.5 h 行车数据,共 12 960 帧数据,原文件大小为 70 787 KiB。不同压缩算法验证结果如表 2 所示。

从表 2 可看出:在压缩率方面,LZW 压缩算法优于改进的 RLE 压缩算法,但消耗时间是改进的 RLE 压缩算法的 62 倍。因解压缩时间与压缩时间

成正比,LZW 压缩算法消耗时间较长,不适合后续的大批量数据分析使用。经综合对比,采用本文改进的 RLE 压缩算法压缩关键帧数据是有优势的,在压缩速度快的基础上还可获得较强的压缩能力。

表 2 不同压缩算法验证结果

Tab. 2 Verification results with different compression algorithms

算法	压缩后数据大小/KiB	消耗时间/s	压缩率/%
LZW 压缩算法	27 307	130.547	38.58
改进的 RLE 压缩算法	37 912	2.113	53.56

## 2.3 降低数据整体冗余

通过上述两种压缩算法,数据被写入文件或用于无线传输,但为了能存储更长时间的数据且加快数据下载的速度,可对写入文件的数据进一步压缩。设计每小时记录 1 个独立文件,当记录文件更换后,可对记录完成的文件进行压缩。记录后文件的压缩不影响新数据处理的实时性,对压缩实时性要求不高,工程上通常选择成熟的压缩库(.zip、.tgz 等格式)进行操作。本文选择 ZIP 压缩算法对文件整体进一步压缩,作为前两种压缩算法的补充。

## 2.4 数据多维度实时压缩方案

本文提出了数据多维度实时压缩方案,其压缩流程如图 2 所示。

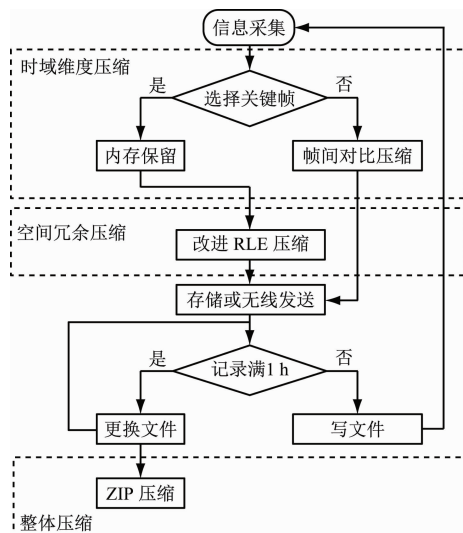


图 2 数据多维度实时压缩方案的压缩流程

Fig. 2 Compression flow chart of the multi-dimensional real-time compression scheme

## 3 试验与应用

### 3.1 算法实时性测试

本文实时性指的是压缩时间满足数据记录或



传输周期。工程上,数据记录周期常用 20 ~ 100 ms,无线传输周期常用 500 ~ 1 000 ms,考虑严酷条件,以 20 ms 为数据记录及传输周期。某项目每周约有 6 000 B 的数据量,实测数据写入 flash 及无线传输时间不超过 5 ms,再预留 5 ms 裕量,则额外引入的压缩工作不超过 10 ms 都足以满足列车数据记录和传输的实时性要求。

在上述多维度实时压缩方案中,ZIP 压缩是对已完成记录文件的压缩,因此不考虑 ZIP 压缩实时性要求。在数据实时记录或传输过程中,对于每周数据,关键帧使用 RLE 压缩算法,非关键帧使用帧间压缩算法。上述两种算法都可以通过对数据帧一次循环遍历即可完成,算法为线性复杂度  $O(N)$  ( $N$  为数据规模)。

使用 2 万帧实际行车数据,对每帧数据采用两种算法进行压缩,消耗时间如表 3 所示。

表 3 两种压缩算法消耗时间验证结果

Tab. 3 Time-consuming verification results with two compression algorithms

算法	平均压缩消耗时间/ms
帧间压缩算法	0.04
改进的 RLE 压缩算法	0.16

由表 3 可见:添加数据压缩后,1 帧数据处理时间增量不超过 0.16 ms,远低于 10.00 ms 预期,满足实时性要求。

### 3.2 实际行车数据验证

采用某城市轨道交通线路行车数据,选取 4 个不同日期下的行车数据(含检修、运行等工况)进行算法验证。

以 A、B、C、D 代表 4 份数据,压缩前数据大小分别为 1 022.4 MiB、750.0 MiB、901.0 MiB、732.0 MiB。试验采用本文提出的压缩方案及几种对比验证方案,结果如表 4 所示。

由表 4 可以看出:本文所研究压缩方案(帧间压缩 + 改进的 RLE 压缩 + ZIP 压缩)的平均压缩率最小,为 3.35%,压缩能力最强;原始文件数据直接采用 ZIP 压缩方案下的平均压缩率为 4.20%,使用 LZW 压缩算法对每帧数据压缩只能获得 37.77% 的平均压缩率,即使 LZW 压缩 + ZIP 压缩方案也仅能获得 19.44% 的平均压缩率。需要指出的是,更低的压缩率并非本文所追求,磁盘数据写入量或无线数据传输量降低才是本文更为关注的重点。对

比数据直接写入磁盘,本文方案磁盘数据写入量或车地数据传输量缩减至 11.83%。虽然从压缩率角度来看,本文方案 3.35% 的平均压缩率和 ZIP 压缩方案 4.20% 的平均压缩率相差不大,但大幅降低了磁盘擦写次数和无线传输带宽占用,对磁盘寿命的延长、磁盘故障率的降低及无线通信带宽的降低均有益处。

表 4 多维度实时压缩方案验证结果

Tab. 4 Verification results of the multi-dimensional real-time compression scheme

方案	不同数据压缩后大小/MiB				平均压缩率/%
	数据 A	数据 B	数据 C	数据 D	
帧间压缩 + 改进的 RLE 压缩	113.0	84.8	111.0	92.6	11.83
帧间压缩 + 改进的 RLE 压缩 + ZIP 压缩	31.9	23.6	31.9	26.4	3.35
ZIP 压缩	41.1	30.1	39.3	32.2	4.20
LZW 压缩	374.0	282.0	342.0	285.0	37.77
LZW 压缩 + ZIP 压缩	198.0	146.0	176.0	142.0	19.44

### 3.3 方案应用

将本项目方案应用于某城市轨道交通线路的网络控制系统数据记录和发送单元。使用 100 ms 周期采集约 5 500 B 的数据,存储周期为 30 d,采用 8 GiB flash 磁盘存储,可存储数据的空间约 6.5 GiB。

实测每列列车行车数据为 2.0 ~ 3.5 GiB/d。若不对行车数据进行压缩,磁盘约 2 d 擦写一遍。使用本文方案,对每帧数据进行帧间压缩或 RLE 压缩并写入磁盘后,每 0.5 h 对文件进行一次 ZIP 压缩,压缩后的数据约 75 ~ 130 MiB/d,足够满足 30 d 数据量存储的要求。

更为关键的是,以往项目频繁遭遇磁盘存储故障和数据损坏问题,但本项目由于写入磁盘的数据量大幅减少,已经连续运行 1 年多都未出现磁盘或数据损坏的故障,从而有效解决了长期困扰项目团队的难题。

## 4 结语

1) 通过有效压缩行车数据,可以在相同的硬件成本下存储更长时间段的数据;有效减少磁盘写入量,降低磁盘擦写次数,从而降低磁盘故障率,提高存储卡的使用寿命;有效减少无线传输数据量,降低传输带宽占用和无线通信费用。

2) 本文在工程经验的基础上研究了适用于城市轨道交通行车数据的压缩方法,实际应用也验证了方案的有效性,可为行车数据压缩工程应用提供参考。

## 参考文献

- [1] 余祖俊, 艾丽, 史红梅. 车载数据记录压缩算法的研究[J]. 电子产品世界, 2001, 8(7): 12.  
YU Zujun, AI Li, SHI Hongmei. Research on compression algorithm of the data for ongoing locomotive[J]. Electronic & Computer Design World, 2001, 8(7): 12.
- [2] 白菊兰. 城市轨道交通车辆信息远程无线下载系统[J]. 铁道车辆, 2018, 56(9): 24.  
BAI Julan. The remote radio unloading system for urban rail traffic information[J]. Rolling Stock, 2018, 56(9): 24.
- [3] 李彬. 城市轨道交通车辆数据无线下载装置研发[J]. 城市轨道交通研究, 2017, 20(2): 94.  
LI Bin. Research and development of wireless transmission system for urban rail transit vehicle[J]. Urban Mass Transit, 2017, 20(2): 94.
- [4] 马俊源. 工业物联网高速大容量在线数据分析记录仪设计与应用[D]. 北京: 北方工业大学, 2021.  
MA Junyuan. Design and application of high-speed and large-capacity online data analysis recorder for industrial internet[D]. Beijing: North China University of Technology, 2021.

(上接第 252 页)

## 参考文献

- [1] 余浩伟, 徐银光, 李涛, 等. 悬挂式单轨交通在国内的适应性研究[J]. 铁道工程学报, 2019, 36(4): 70.  
YU Haowei, XU Yinguang, LI Tao, et al. Research on the adaptability of suspended monorail in China[J]. Journal of Railway Engineering Society, 2019, 36(4): 70.
- [2] 王玉华, 王孟君, 谢宇杰. 悬挂式单轨系统结构方案研究[J]. 科学技术创新, 2022(6): 147.  
WANG Yuhua, WANG Mengjun, XIE Yujie. Research on the structure scheme of suspended monorail system[J]. Scientific and Technological Innovation, 2022(6): 147.
- [3] 杨翠屏, 张津, 邢扬. 悬挂式单轨轨道梁制造工艺研究[J]. 城市轨道交通研究, 2019, 22(4): 48.  
YANG Cuiping, ZHANG Jin, XING Yang. Research on the manufacturing process of suspended monorail beam[J]. Urban Mass Transit, 2019, 22(4): 48.
- [4] 国家铁路局. 铁路桥梁钢结构设计规范: TB 10091—2017[S]. 北京: 中国铁道出版社, 2017.  
National Railway Administration of the People's Republic of China. Code for design of steel structure of railway bridge: TB 10091—2017[S]. Beijing: China Railway Publishing House, 2017.

- [5] 邢旺, 郁聪冲. 数据记录器中 NAND Flash 的坏块管理方法[J]. 舰船电子工程, 2016, 36(8): 132.  
XING Wang, YU Congchong. Bad block management methods of NAND flash memory in data recorder[J]. Ship Electronic Engineering, 2016, 36(8): 132.
- [6] 刘敏丽. 航天遥测数据无损压缩技术研究与应用[D]. 北京: 北京理工大学, 2016.  
LIU Minli. Research and Implementation of Lossless Compression Technique for Space Telemetry Data[D]. Beijing: Beijing Institute of Technology, 2016.
- [7] 王平. LZW 无损压缩算法的实现与研究[J]. 计算机工程, 2002, 28(7): 98.  
WANG Ping. Realization and research of LZW lossless compression algorithm[J]. Computer Engineering, 2002, 28(7): 98.
- [8] 全立新. Run-Length 压缩算法的改进及 C++ 的实现[J]. 长沙电力学院学报(自然科学版), 2004, 19(2): 29.  
QUAN Lixin. The improvement of Run-Length compression algorithms and its execution of C++[J]. Journal of Changsha University of Electric Power (Natural Science), 2004, 19(2): 29.

· 收稿日期:2022-08-03 修回日期:2022-10-17 出版日期:2025-01-10

Received:2022-08-03 Revised:2022-10-17 Published:2025-01-10

· 通信作者:李德祥,高级工程师,ldxab@163.com

· ©《城市轨道交通研究》杂志社,开放获取 CC BY-NC-ND 协议

© Urban Mass Transit Magazine Press. This is an open access article under the CC BY-NC-ND license

- [5] 中国国家铁路集团有限公司. 铁路钢桥制造规范: Q/CR 9211—2015[S]. 北京: 中国铁道出版社, 2015.  
China State Railway Group Co., Ltd. Code for fabrication of railway steel bridge: Q/CR 9211—2015[S]. Beijing: China Railway Publishing House, 2015.
- [6] 李伟. 鱼腹型变截面连续钢箱梁“卧装反造法”3+1 模式施工技术[J]. 市政技术, 2021, 39(2): 40.  
LI Wei. Construction technology of model 3+1 applied in installing of continuous steel box girder with fish-bellied variable cross section by horizontal installation and reverse construction method[J]. Municipal Engineering Technology, 2021, 39(2): 40.
- [7] 胡广瑞. 大型公路钢箱梁整体拼装制造线形和尺寸的控制[J]. 钢结构, 2006, 21(5): 74.  
HU Guangrui. The methods of profile and dimension control during assembly of large steel box girder for highway[J]. Steel Construction, 2006, 21(5): 74.

· 收稿日期:2022-09-26 修回日期:2022-10-21 出版日期:2025-01-10

Received:2022-09-26 Revised:2022-10-21 Published:2025-01-10

· 第一作者:尹桂芳,工程师,125311264@qq.com

通信作者:龚文强,工程师,2315666282@qq.com

· ©《城市轨道交通研究》杂志社,开放获取 CC BY-NC-ND 协议

© Urban Mass Transit Magazine Press. This is an open access article under the CC BY-NC-ND license