

卫星遥测数据实时压缩算法设计与实现

史礼婷¹, 张 骞², 程科文¹, 吕海全¹

(1 航天行云科技有限公司 武汉 430040;

2 宁波天擎航天科技有限公司 宁波 315100)

摘要: 为了满足多星并行遥测数据处理和海量数据高并发分析的性能需求, 便于后续进行数据挖掘、智能预警, 本文提出并实现了一种卫星遥测数据实时压缩算法。针对遥测数据的特点, 提出了遥测自适应分类方法, 采用改进型 RLE (Run Length Encoding, 行程编码) 压缩和增量压缩结合的算法, 结合数据库技术, 实现了遥测数据的压缩。在某型号卫星研制项目中, 采用了该算法进行数据压缩, 统计分析表明: 该算法起到了很好的压缩效果。

关键词: 遥测数据; 压缩算法; 自适应分类; 改进型 RLE; 增量压缩

中图分类号: TP301.6; V557+.3

文献标志码: A

文章编号: 2095-1000(2023)05-0022-06

DOI: 10.12347/j.ycyk.20230510002

引用格式: 史礼婷, 张骞, 程科文, 等. 卫星遥测数据实时压缩算法设计与实现[J]. 遥测遥控, 2023, 44(5): 22–27.

Design and implementation of real-time compression algorithm on satellite telemetry data

SHI Liting¹, ZHANG Qian², CHENG Kewen¹, LYU Haiquan¹

(1. Leobit Technology Co., Ltd., Wuhan 430040, China;

2. Ningbo Tianqing Aerospace Technology Co., Ltd., Ningbo 315100, China)

Abstract: In order to meet the requirements of multi-satellite parallel telemetry data processing and high-concurrency analysis for massive data, facilitate data mining and intelligent warning, this paper proposes and implements a real-time compression algorithm for satellite telemetry data. Based on the characteristics of telemetry data, a telemetry adaptive classification method is proposed, using an improved RLE compression and incremental compression algorithms combined with database technology to achieve compression of telemetry data. In the development project of a certain type of satellite, this algorithm was used for data compression, and statistical analysis shows that this algorithm has achieved good compression results.

Keywords: Telemetry data; Compression algorithm; Adaptive classification; Improved RLE; Incremental compression

Citation: SHI Liting, ZHANG Qian, CHENG Kewen, et al. Design and implementation of real-time compression algorithm on satellite telemetry data[J]. Journal of Telemetry, Tracking and Command, 2023, 44(5): 22–27.

引 言

近年来, 商业航天呈现出爆发式增长。新一代“千星座”乃至“万星座”的空间星座系统启动建设^[1], 以 OneWeb、Starlink 等为代表的商业航天公司, 均推出了巨型星座计划^[2,3], 利用卫星为社会提供更加精准、高效的信息通信服务。同时, 随着运载能力不断提升, 卫星发射越来越密集。2021 年 5 月, 猎鹰-9 火箭首次实现“一箭十发”, 迈过火箭重复利用常态化的关键一步, 给全

球航天发射市场带来巨大冲击, 而随着“星链”卫星的密集发射, 在人类历史上首次实现千颗以上规模卫星的快速部署^[4]。

随着卫星数量的不断增加, 卫星测试技术也迎来了新挑战, 其中最为重要的两点是: 一, 卫星由以往的单颗卫星定制研制模式逐渐转变为模块化、批量化研制; 二, 卫星测试更为趋向于自动化、通用化、批量化、并行化^[5]。面对数以千计的卫星以及不断增长的遥测数据流, 传统的数据处理方案难以应对。不但要满足多卫星并行批产

测试、发射、在轨运营的基础要求,而且要满足数据实时分析、预警等高性能使用需求。

卫星遥测时序数据是以数据流的形式呈现的,具有传输度快、数据量大和数据维度高等特点^[6],这些特征给实时数据处理和传输带来了非常大的挑战。如何高效存储、处理和响应这些数据流,以满足实时监控、快速反应和预警需求是地面站面临的难题。

因此,在面临卫星遥测数据的实时处理时,需要考虑如何实现高效的数据存储、实时处理和即时响应的策略。其中,卫星遥测数据实时压缩是提高实时数据处理性能的一种有效方法,它可以降低信息传输速率、降低通信信道带宽,降低对数据存储介质的要求^[7,8],便于后续进行数据挖掘、预警,大大提高了海量遥测数据实时处理性能。

基于以上应用需求,本文针对遥测数据的时序性、数量大和种类多的特点,提出了一种遥测数据实时压缩处理算法。在遥测自适应分类处理的基础上,采用RLE压缩和增量式压缩结合的方法,结合了数据库存储/缓存技术,在某卫星研制项目中,验证该算法的数据压缩性能。

1 遥测数据特点

1.1 时间序列

卫星在运行过程中,星载计算机按周期采集卫星内部各个单机或分系统的运行状态数据,按CCSDS(空间数据系统咨询委员会)格式进行分包存储^[9],形成遥测数据。在轨运行阶段,卫星会在经过地面站测控弧度期间,进行遥测数据下发;而地面测试阶段,除了无线信道传输之外,还会通过星地通道实时下发CCSDS遥测数据,便于研发人员对卫星状态进行实时监控。无论在哪个阶段,遥测数据均为基于时间序列生成的结构化数据。

1.2 数据量大、种类多

随着运载能力的增强、卫星星座应用需求的增加,对多颗卫星的并行处理的性能要求日益提高。与此同时,单通道无线数据传输能力也在逐渐增强,地面需要实时处理的遥测数据越来越庞大。在地面测试阶段,一天下发遥测数据量级达GB甚至TB。

卫星的遥测参数可能多达上千个,参数种类也很多,系统复杂;卫星飞行空域广,通信距离

远,加之遥测参数的格式编排复杂,这样就对遥测系统的数据采集、传输、解调提出了很高的要求^[10]。

2 遥测数据压缩

根据信息论的基本概念,数据压缩技术可以分为无损数据压缩和有损数据压缩两大类^[11]。在航天遥测系统中,遥测数据对于航天器飞行轨道的确定、故障的定位处理具有重要意义,地面测控站需要接收完整准确的航天遥测数据^[12],因此,在航天遥测系统中,采用无损数据压缩算法对航天遥测数据进行处理。

常规的无损压缩算法包括霍夫曼编码、行程编码、算术编码和LZ系列编码。其中,霍夫曼编码和算术编码都是基于统计模型的压缩编码方法,LZ系列编码是基于字典的压缩编码方法^[13,14]。目前,已有很多对压缩技术的应用研究。文献[12]采用改进字典搜索方式的LZW算法(串表压缩算法)和RLE算法(游程编码算法)结合,采用FPGA实现遥测数据无损压缩。文献[15]将RLE算法与LZW算法相结合,并且提出删除使用频率较小的字典分支对字典进行更新。文献[16]提出融合LZW与霍夫曼编码的联合压缩算法,即先对数据进行LZW压缩,再进行霍夫曼编码,仿真结果验证了改进算法的有效性。文献[17]对LZW算法的FPGA硬件实现方案进行研究,将LZW算法的字典划分为多个小字典,并且在硬件实现时采用并行搜索方式。文献[18]提出了一种航天器分包遥测压缩算法,将RLE应用到分包遥测数据压缩中。上述文献均采用多种压缩编码方式结合的方式实现数据压缩,大多应用于星载遥测数据压缩,对于地面遥测数据的压缩处理较少涉及。

为达到实时快速压缩的效果,本文以经典的无损压缩算法——RLE编码思想为基础,结合遥测数据的特点和数据库技术,提出了遥测自适应分类方法,采用改进型RLE压缩算法和增量压缩结合的方法实现遥测数据的压缩。

2.1 遥测压缩方法设计

行程编码又被称为游程编码,行程长度(RL, Run Length)是指字符(或者信号采样值)构成的数据流中各字符重复出现而形成的字符串的长度^[19]。其基本思想是对于连续出现 y 次的重复数据 b ,使用 yb 来表示^[19],例如:aaaabccccdddee,则可用

4a1b-5c3d2e 来代替。该算法思想简单, 压缩速度快, 但是对于重复次数大于 2 的数据才能起到很好的压缩效果, 例如: 完全不重复的数据 abcde, 用 1a1b-1c1e 来代替反而会造成数据膨胀。因此, 在应用过程中将 RLE 算法进行改进, 对于频繁变化的数据, 不再记录重复长度, 从而避免数据膨胀。

通过对卫星遥测数据进行分析发现, 数据主要分为二类: 一是状态量, 在固定的数值内变化, 如记录开/关、运行模式、存储区编号等; 二是数值型, 数值变化范围较广, 如电流、电压、温度等需要公式计算的数值, 或是版本号、指令码、计数、地址等十进制/二进制/十六进制。在这些遥测数据中, 有些会随着时间的缓慢变化, 如电流、电压、温度等测量值; 而有些数据, 在短期内, 会处于相对固定的数值。

在设计遥测数据压缩算法时, 应充分考虑遥测数据的特点。本文提出的遥测数据压缩处理算法数据流如图 1 所示, 包括数据接收、数据预处理、数据自适应分类、数据压缩、数据存储和数据查询分析几个步骤。

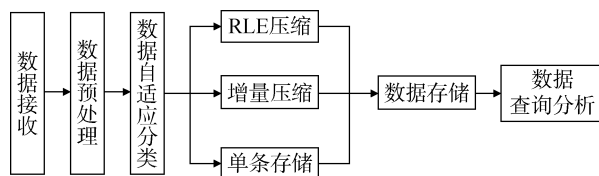


图 1 遥测数据压缩处理数据流

Fig. 1 Data flow diagram for telemetry data compression processing

其中, 遥测自适应分类和数据压缩为压缩算法核心, 具体说明如下:

① 自适应分类: 数据挖掘分类方法有决策树分类、贝叶斯分类、神经网络分类、支持向量机分类^[20]。本文采用决策树分类思想, 建立决策树模型如图 2 所示, 根据数值变化偏移距离进行分类。首先建立数据临时字典, 存放历史数据集, 通过对实时数值和历史数值进行比较, 得到数据偏移值。根据偏移值进行自适应分类, 若偏移量为 0, 则为重复数据, 放入改进型 RLE 压缩处理类中; 若偏移量不等于 0, 则决策是否为状态量, 若为状态量, 则放入增量压缩处理类中, 不为状态量, 则放入其他遥测处理类中。

② 遥测数据主要分为三类进行数据压缩:

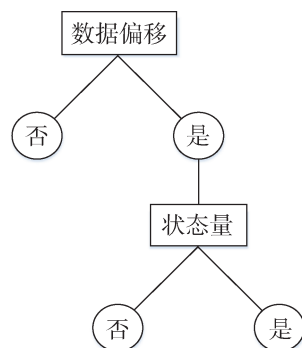


图 2 自适应分类决策树

Fig. 2 Adaptive classification decision tree

(1) 改进型 RLE 压缩: 按 RLE 算法思想进行快速压缩, 根据时间序列特点, 算法中的行程长度 L 选定为时间窗口, 重复数据按时间窗口压缩成一条数据存储。

(2) 增量压缩: 由于此类数据数值随时间不断变化, 不宜采用 RLE 压缩处理。但是此类数据只在固定数值范围内变化, 针对数值变化特点, 采用时间增量式压缩方法, 记录各个状态量增量时间的方式进行数据压缩, 将多个时间点的数据压缩成一条数据进行记录。

(3) 其他遥测量: 按数据产生时间对单条数据进行处理, 避免了 RLE 算法导致的数据膨胀问题。

2.2 压缩算法实现

为保障实时处理的性能, 采用多线程多级缓存并行处理的思路进行算法实现, 对遥测数据的压缩处理包括数据预处理、数据解析、数据自适应分类这几个步骤, 流程图如图 3 所示。图中的相关步骤均采用独立线程进行处理, 充分利用数据队列和 Redis 数据库进行缓存, 提高算法处理性能。

压缩算法流程具体说明如下:

① 数据预处理: 为实现多星数据并行处理, 地面需要部署多个数据监听节点, 通过无线/有线链路多路同时进行数据接收。接收到的数据信号经过滤波、放大、解调等处理后, 需要经过数据预处理, 才能有效地剔除因噪声、干扰或信号衰减而引起的误码。预处理主要通过数据解扰和 CRC 校验(循环冗余校验)实现。数据加解扰有助于保障航天系统的安全和保密性, 系统采用密钥解密算法进行数据解扰后, 利用多项式对数据进行特定算法运算, 产生 CRC 校验码, 比较 CRC 校验码的正确性即可快速检验数据的正确性。

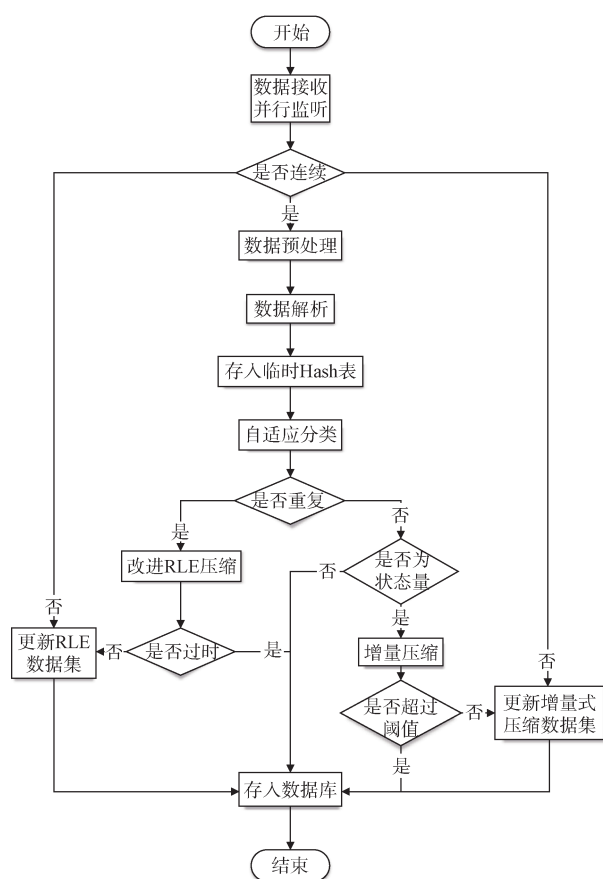


图3 压缩算法流程图

Fig. 3 Compression algorithm flowchart

② 数据解析：通过数据预处理之后的数据原码，将实时放入数据解析线程。数据解析线程根据已配置好的CCSDS遥测协议，对数据实时进行工程值解析。解析之后的数据，是带有时间标记和解析方式分类的数据工程值，以Key-Value形式存在内存中，形成临时Hash数据表，其中Key以遥测代号作为关键字，Value记录数据起始UTC时间和工程值。

③ 自适应分类：接收到实时遥测解析数据集后，需要查询历史遥测Hash表中对应数据，实时计算数据偏移量，根据决策树模型进行自动分类，若偏移量为0，则表明数值不变，采用RLE压缩方法，进入RLE压缩队列；若偏移量不为0，则判断是否为状态量，若为状态量，则采用时间增量式压缩方法，进入增量压缩队列，若不为状态量，则直接进入数据持久化队列。

④ 改进型RLE压缩：结合遥测数据的时序特点，将RLE算法思想中的长度定为数据的持续时间，建立压缩数据集，令 t_0 为起始时间， t_1 为结束

时间，则数据记录为 (t_0, t_1, value) 。在接收到分类数据之后，根据遥测代号找到压缩数据集中对应的数据，更新数据集的数据结束时间；若RLE数据集内部，有结束时间小于 t_1 的记录，则认为该数据已过时，数据不再连续重复，则需要将数据加入数据持久化队列。

⑤ 增量式压缩：根据遥测代号与现有数据集对应数据进行比较，若现有数据没有起始时间，即为第一次进入数据集数据，将以此时的UTC时间作为起始时间；若有起始时间，则记录该工程值距离起始时间的时间增量，令 t_0 为起始时间， Δt 为时间增量，则数据记录为 $(t_0, \Delta t_1, \Delta t_2, \dots, \Delta t_n, \text{value})$ 。为保障数据量均衡，设置时间增量阈值600(该数值可调)，若 Δt 超过600，则直接进入数据持久化队列。

⑥ 数据持久化：本文采用MySQL对解析之后的数据进行持久化。接收待入库的数据后，根据数据来源打上压缩方法标记，存入数据库中。为防止频繁update数据库，对系统性能造成消耗，采用Redis内存数据库对数据进行缓存，同时，对数据接收时间进行监听，若接收时间间隔大于2秒，则认为数据中断，将现有数据集批量存入数据库。

2.3 压缩数据还原

在卫星实际测试中，需要随时对数据进行图形或数值的查看，压缩的数据还原是必不可少的环节。在数据压缩的过程中，自适应分成RLE压缩和增量式压缩两类进行数据处理，因此，在数据还原的时候，也要根据压缩方法标记进行分类归一化处理，具体说明如下：

① 数据归一化原则：根据遥测数据的特点，还原后的遥测数据至少应包含时间、工程值、遥测处理方式这几个要素。

② RLE压缩数据还原：RLE压缩后的数据，包含数据起止时间和工程值，因此，根据数据起止时间进行数据周期插入即可对数据进行还原。

③ 增量式压缩数据还原：根据状态量起始时间和时间增量，以 $t=t_0+\Delta t$ ，进行实际时间的计算，即可对数据进行还原。

3 性能分析

本文采用Java作为后台编程语言，实现数据压缩和还原算法的核心代码。在某卫星研制项目中，对该算法进行了实际部署和效果验证。

遥测数据采用数据库进行结构化存储, 设单条记录占用的空间为 m 字节, 压缩前数据总数量为 N_1 , 压缩之后的数据总数量为 N_2 , 则数据压缩数量为 $N_1 - N_2$, 压缩率为 $(N_1 - N_2) \times m / (N_1 \times m) \times 100\% = (N_1 - N_2) / N_1 \times 100\%$ 。

为了统计压缩率, 同时保存了压缩前和压缩后的数据, 进行数据对比和验证。本文选取了某月的 8 日、9 日、10 日连续三天的测试数据, 对数据总量进行了分类统计, 见表 1、表 2、表 3。统计表明: 三天的压缩率分别为 88.27%、88.10%、90.42%, 压缩率达到了 88% 以上, 其中, 状态量和部分变化较小的数值型遥测数据压缩率达到了 99% 以上。

表 1 8 日数据压缩量统计

Table 1 Statistics of data compression on the 8th

数据类型	压缩前 (条)	压缩后 (条)	压缩量 (条)	压缩率
状态量				
十六进制	2 394 174	2 433	2 391 741	99.90%
十进制	1 689 373	1 752	1 687 621	99.90%
二进制	6 410 226	8 498	6 401 728	99.87%
二进制	437 755	346	437 409	99.92%
十六进制	2 602 505	25 819	2 576 686	99.01%
十进制	20 337 140	255 312	20 081 828	98.74%
数值型				
32 位双精度浮点数	250 288	46 752	203 536	81.32%
公式计算值	7 496 288	2 794 080	4 702 208	62.73%
电压值	3 627 436	2 040 666	1 586 770	43.74%
范围	281 439	162 793	118 646	42.16%
合计	45 526 624	5 338 451	40 188 173	88.27%

进一步对遥测数据的自适应分类效果进行统计分析, 由于压缩总量由改进型 RLE 压缩和增量压缩组成, 若不采用自适应分类算法, 则全部采用改进型 RLE 压缩算法进行数据压缩。因此, 对压缩总量和增量压缩量进行统计, 增量压缩量的占比即为采用自适应分类算法后压缩提高的比率。统计信息如表 4 所示, 可以看到, 采用自适应分类进行压缩后, 8 日、9 日、10 日的压缩率分别提高了 33.55%、33.50%、32.75%。

为验证数据还原的效果, 对压缩前和压缩后的数据分别进行了历史数据查询图形对比显示, 如图 4 所示, 图中横轴为卫星地面测试的时间, 纵轴为遥测数据的数值, 可以看出, 压缩数据还原

表 2 9 日数据压缩量统计

Table 2 Statistics of data compression on the 9th

数据类型	压缩前 (条)	压缩后 (条)	压缩量 (条)	压缩率
状态量				
十六进制	2 011 651	2 564	2 009 087	99.87%
十进制	1 395 154	1 850	1 393 304	99.87%
二进制	5 497 912	7 296	5 490 616	99.87%
二进制	375 455	432	375 023	99.88%
十六进制	2 226 108	21 840	2 204 268	99.02%
十进制	17 377 667	183 782	17 193 885	98.94%
数值型				
32 位双精度浮点数	214 688	64 595	150 093	69.91%
公式计算值	6 357 114	2 477 636	3 879 478	61.03%
电压值	3 111 236	1 718 706	1 392 530	44.76%
范围	241 389	138 081	103 308	42.80%
合计	38 808 374	4 616 782	34 191 592	88.10%

表 3 10 日数据压缩量统计

Table 3 Statistics of data compression on the 10th

数据类型	压缩前 (条)	压缩后 (条)	压缩量 (条)	压缩率
状态量				
十六进制	1 825 730	3 969	1 821 761	99.78%
十进制	1 266 688	3 266	1 263 422	99.74%
二进制	4 994 177	11 284	4 982 893	99.77%
二进制	341 050	660	340 390	99.81%
十六进制	2 023 718	6 298	2 017 420	99.69%
十进制	15 797 201	137 848	15 659 353	99.13%
数值型				
32 位双精度浮点数	195 024	29 879	165 145	84.68%
公式计算值	5 924 680	1 712 133	4 212 547	71.10%
电压值	2 826 108	1 365 706	1 460 402	51.68%
范围	219 267	122 354	96 913	44.20%
合计	35 413 643	3 393 397	32 020 246	90.42%

表 4 自适应算法效果统计

Table 4 Adaptive algorithm performance statistics

时间	压缩总量(条)	增量压缩量(条)	占比
8 日	40 188 173	13 483 967	33.55%
9 日	34 191 592	11 521 513	33.50%
10 日	32 020 246	10 487 559	32.75%

后, 与压缩前完全一致。分析表明: 本文提出的遥测压缩算法, 实现了遥测数据的压缩和还原, 已经在工程实践中运用, 达到了比较好的压缩效果。

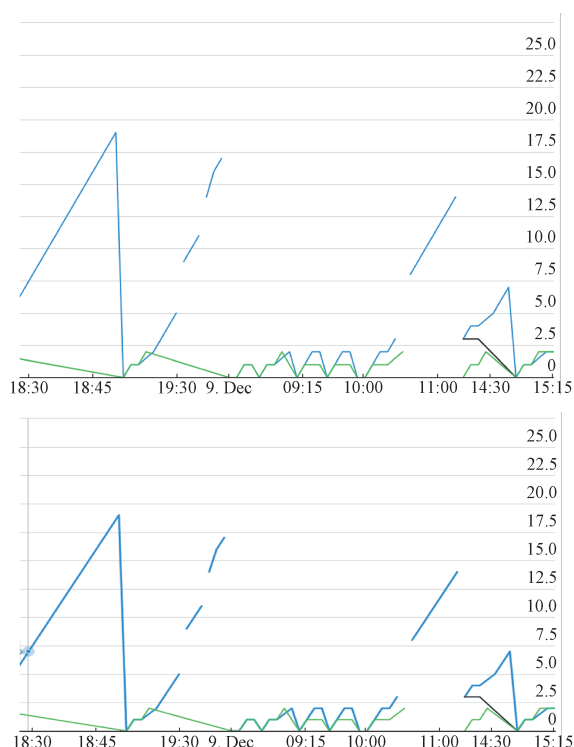


图 4 压缩前(上)和压缩后(下)数据查询对比图(截图)

Fig. 4 Comparison of data queries before compression(up) and after compression(down)

4 结束语

经过实际验证, 本文提出并实现的遥测自适应分类方法, 以及在此基础上进行的 RLE 压缩和增量压缩结合的算法, 达到了较好的数据压缩效果, 为遥测数据分析、预警打下了良好的基础。后续还可以在此基础上进一步研究, 对缓变非状态量的遥测数据进行研究分析, 采用合适的算法对其进行数据压缩, 进一步提高数据压缩率。

参考文献

- [1] 李峰, 禹航, 丁睿, 等. 我国空间互联网星座系统发展战略研究[J]. 中国工程科学. 2021, 23(4): 137-144.
LI Feng, YU Hang, DING Rui, et al. Development strategy of space internet constellation system in China[J]. Engineering Sciences, 2021, 23(4): 137-144.
- [2] ALBULET M. SpaceX non-geostationary satellite system: Technical information to supplement schedules[EB/OL]. (2016-11-15) [2023-05-10]. <https://cdn.geekwire.com/wp-content/uploads/2016/11/Technical-Attachment.pdf>.
- [3] PORTILLO I D, CAMERON B G, CRAWLEY E F, et al. A technical comparison of three low earth orbit satellite constellation systems to provide global broadband[EB/OL]. (2018-10-01) [2023-05-10]. <https://pdfs.semanticscholar.org/a2b7/931f4e93c6044d5cc4af78d6c60b1db6d-bc3.pdf>.

- [4] 马良, 何建伟, 王浩铭. 全球商业航天发展新趋势及思考[J]. 国际太空, 2021(8): 25-30.
- [5] 张云鹏. 批量微纳卫星地面并行自动化测试应用技术研究[D]. 哈尔滨: 哈尔滨工业大学, 2021.
- [6] 周忠玉. 遥测时序数据稀有模式挖掘及应用[D]. 南京: 南京航空航天大学, 2020.
- [7] 王继东. 星载遥感图像实时压缩中的关键技术研究[D]. 长沙: 国防科学技术大学, 2007.
- [8] CHEN Z, TU G F, ZHANG C, et al. Design of CCSDS image compression system based on FPGA[J]. Journal of the Graduate School of the Chinese Academy of Sciences, 2011, 28(1): 101-107.
- [9] CCSDS. CCSDS 120.0-G-2 Lossless data compression [S]. Washington D.C.: CCSDS, 2006.
- [10] 孙浩, 于斌. 基于遥测数据的压缩算法设计[J]. 舰船电子工程, 2010, 30(8): 57-59.
SUN Hao, YU Bin. Design of compression algorithm based on telemetry data[J]. Ship Electronic Engineering, 2010, 30(8): 57-59.
- [11] SALOMON D. Data compression: The complete reference[M]. New York: Verlag New York, Inc., 2000.
- [12] 刘敏丽. 航天遥测数据无损压缩技术研究与实现[D]. 北京: 北京理工大学, 2016.
- [13] TAMM U. Communication complexity and data compression[C]// Information Theory and Applications Workshop (ITA), 2013. IEEE, 2013: 1-4.
- [14] KARTHICK K, KUMAR R V. An energy-saving routing algorithm for wireless body sensor network using data compression technique[J]. International Journal of Computer Applications, 2015, 114(15): 26-29.
- [15] SUN L, CHEN B, NAN S, et al. A dictionary updating scheme incorporating words elimination into Quantized Kernel Least-Mean-Squares for changing environments [C]// Digital Signal Processing (DSP), 2015 IEEE International Conference on. IEEE, 2015.
- [16] ZHANG Y, LI X, HUA D, et al. An Lidar data compression method based on improved LZW and Huffman algorithm[C]//Electronics and Information Engineering (ICEIE), 2010 International Conference On. IEEE, 2010: 250-254.
- [17] EDWARDS J A, VISHKIN U. Parallel algorithms for Burrows-Wheeler compression and decompression[J]. Theoretical Computer Science, 2014, 525(4): 10-22.
- [18] 闫国瑞, 李志刚, 史简, 等. 应用行程编码的航天器分包遥测压缩算法研究[J]. 航天器工程, 2015, 24(2): 58-61.
YAN Guorui, LI Zhigang, SHI Jian, et al. Research on packet telemetry compression algorithm using RLE[J]. Spacecraft Engineering, 2015, 24(2): 58-61.
- [19] NISHITSUJI T, SHIMOBABA T, KAKUE T, et al. Fast calculation of computer-generated hologram using run-length encoding based recurrence relation[J]. Optics Express, 2015, 23(8): 9852-9857.
- [20] 郭炜星. 数据挖掘分类算法研究[D]. 杭州: 浙江大学, 2008.

[作者简介]

史礼婷 1984年生, 硕士, 高级工程师。
张 骞 1984年生, 硕士, 高级工程师。
程科文 1989年生, 硕士, 工程师。
吕海全 1993年生, 硕士, 工程师。

(本文编辑: 傅 杰)