

基于边缘计算的数据无损压缩方法

张小梅, 曹莹, 姜平, 江雪梅, 严俊伟, 李达

引用本文

张小梅, 曹莹, 姜平, 江雪梅, 严俊伟, 李达[基于边缘计算的数据无损压缩方法](#)[J]. 计算机科学, 2022, 49(11A): 210500195-6.

ZHANG Xiao-mei, CAO Ying, LOU Ping, JIANG Xue-mei, YAN Jun-wei, LI Da. [Lossless Data Compression Method Based on Edge Computing](#) [J]. Computer Science, 2022, 49(11A): 210500195-6.

相似文章推荐 (请使用火狐或 IE 浏览器查看文章)

Similar articles recommended (Please use Firefox or IE to view the article)

[移动边缘计算中任务卸载研究综述](#)

Survey of Research on Task Offloading in Mobile Edge Computing

计算机科学, 2022, 49(11A): 220400161-7. <https://doi.org/10.11896/jsjcx.220400161>

[边缘环境下轨迹预测性感知的在线边缘服务分配](#)

Novel Predictive Approach to Trajectory-aware Online Edge Service Allocation in Edge Environment

计算机科学, 2022, 49(11): 277-283. <https://doi.org/10.11896/jsjcx.211100029>

[基于负载特征的边缘智能系统性能优化](#)

Workload Characteristics Based Performance Optimization for Edge Intelligence

计算机科学, 2022, 49(11): 266-276. <https://doi.org/10.11896/jsjcx.211000067>

[无人机边缘计算中的资源管理优化研究综述](#)

Survey of Resource Management Optimization of UAV Edge Computing

计算机科学, 2022, 49(11): 234-241. <https://doi.org/10.11896/jsjcx.211100015>

[VEC中基于动态定价的车辆协同计算卸载方案](#)

Dynamic Pricing-based Vehicle Collaborative Computation Offloading Scheme in VEC

计算机科学, 2022, 49(9): 242-248. <https://doi.org/10.11896/jsjcx.210700166>

基于边缘计算的数据无损压缩方法

张小梅^{1,2} 曹 莹¹ 娄 平^{1,2} 江雪梅^{1,2} 严俊伟^{1,2} 李 达^{1,2}

1 武汉理工大学信息工程学院 武汉 430070

2 宽带无线通信和传感器网络湖北省重点实验室 武汉 430070

(may125z@126.com)

摘 要 工业物联网(Industrial Internet of Things, IIoT)可以将各种工业设备、监测仪表以及传感器进行相互连接,设备的运行状态可通过监测仪表与传感器进行全面感知,并根据感知数据对设备状态进行分析与预测。然而要对海量的感知数据进行分析处理需要大量的存储空间与计算能力,将其传送到云平台势必将占用大量的带宽并产生较大的时延,很难满足对设备状态实时分析与诊断的需求。因此,针对工业设备状态感知的数据,提出了基于最优差分 and 线性拟合降熵变换的无损压缩方法,在数据采集的边缘侧对感知数据进行无损压缩,从而大大提高传输效率,使得感知数据能快速传送到云平台进行分析与处理。该方法根据数据差值序列方差大小和采集频率高低从最优差分 and 曲线拟合差值中选取有效的降熵变换进行压缩,并采用 LZO (Lempel-Ziv-Oberhumer) 压缩算法进行二次压缩。在两类不同的数据集上对所提方法进行了测试,实验结果表明,该方法的压缩率最低与最高分别可达 77% 和 93%,同时验证了其无损重构的特性。

关键词: 工业物联网; 边缘计算; 感知数据; 降熵变换; LZO

中图法分类号 TP391

Lossless Data Compression Method Based on Edge Computing

ZHANG Xiao-mei^{1,2}, CAO Ying¹, LOU Ping^{1,2}, JIANG Xue-mei^{1,2}, YAN Jun-wei^{1,2} and LI Da^{1,2}

1 School of Information Engineering, Wuhan University of Technology, Wuhan 430070, China

2 Hubei Key Laboratory of Broadband Wireless Communications and Sensor Networks, Wuhan 430070, China

Abstract Through the Industrial Internet of Things(IIoT), various industrial equipments, monitoring instruments and sensors can be connected to each other. The operating status of equipment can be fully sensed through monitoring instruments and sensors, and the equipment status can be analyzed and predicted based on sensing data. However, the analysis and processing of massive amounts of sensing data requires a lot of storage space and computing power. Sending it to the cloud platform will inevitably take up a lot of bandwidth and cause a large delay. It is difficult to meet the requirements for real-time analysis and diagnosis of device status. Therefore, for the state-aware data of industrial equipment, a lossless compression method based on optimal differential and linear fitting entropy reduction transform is proposed, and the perceptual data is compressed losslessly at the edge of data collection, so that the transmission efficiency is greatly improved and the perceptual data can be quickly transmitted to the cloud platform for analysis and processing. This method selects an effective entropy reduction transform from the optimal difference and curve fitting difference according to the variance of data difference sequence and the acquisition frequency, and uses Lempel-Ziv-Oberhumer(LZO) compression algorithm for secondary compression. The new method is tested on two different data sets. Experimental results show that the lowest and highest compression rate of this method can reach 77% and 93%, respectively. At the same time, the characteristics of its lossless reconstruction are verified.

Keywords Industrial Internet of things, Edge computing, Perceptual data, Reduced entropy transformation, LZO

1 引言

边缘计算是云计算的重要补充,是指在靠近物或数据源的一侧,采用网络、计算、存储、应用核心能力为一体的开放平台,就近提供服务,满足行业业务实时化、应用智能化、传输低

时延等需求。在工业现场,边缘计算技术可以将用不同数据接入协议终端采集的数据转换成统一的语义格式,便于本地计算或传送到云中心。

边缘计算的出现有效地改善了云计算的某些弊端。例如,在万物互联的场景下,需要对海量级别的多源异构数据

基金项目:武汉市科技局应用基础前沿专项(2020010601012176);国家自然科学基金(52075404)

This work was supported by the Application Foundation Frontier Special Project of Wuhan Science and Technology Bureau(2020010601012176) and National Natural Science Foundation of China(52075404).

通信作者:李达(frankli26@whut.edu.cn)

进行融合和实时处理,在云中心和终端间的数据传输由于距离和数据质量问题,造成了很大的带宽资源浪费^[1],并且实时处理效果也很难实现,而且这种长距离传输增加了数据暴露的风险^[2]。边缘计算刚好弥补了这些缺点,但同时也暴露了其他问题。思科在 2016—2021 年的全球云指数中指出:接入互联网的设备数量将从 2016 年的 171 亿增加到 271 亿,每天产生的数据量也在激增,全球设备产生的数据量从 2016 年的 218 ZB 增长到 2021 年的 847 ZB^[3]。数据量规模的增大严重影响需要实时分析处理的需求响应,原因是直接将数据上传至云端,既浪费上传的通信带宽,又浪费计算资源;若是直接存储在边缘服务器中,边缘服务器不管在计算能力还是存储资源方面都远不如云服务器,边缘侧存储容量有限;并且,若要保证数据的有效存储,则边缘侧需要定期删除数据库的数据来存储新数据。因此,在支持小型实时本地数据处理业务的边缘侧设备上增加数据压缩的方案,既可以有效提高边缘侧数据存储效率,又可以减少上传云端时的带宽消耗。

工业现场的机器设备一般是 24 小时不间断运行,不断产生新的数据。机器设备数据就是工业生产线上设备、机器、产品等方面的数据,大多是由传感器、设备仪器仪表进行采集产生,反映生产运行、设备状态、产品质量控制等方面的生产运行状态。这种数据获取方式多样,空间上较分散,数据间关联性强,具有强时间相关性。

边缘服务器具备一定的计算资源,虽然无法承担如云计算般的计算能力,但是可以在数据传送到云之前帮助提高数据分流传输效率,边缘侧可以利用数据分块压缩技术减少数据传输量。文献[4]中提到将数据压缩技术部署在边缘侧的研究目前几乎没有。

所以,在数据量大和边缘服务器计算资源有限的情况,研究数据压缩方法可以有效解决向云端传送数据带宽占用较大和传输时延的问题。边缘计算技术可以解决工业现场的多源数据采集接入问题,在此基础上,利用边缘侧有限的条件实现数据的高性能压缩。本文为处理从工业现场传感器等仪器采集上来的数据,用边缘侧的计算资源对数据进行压缩,以此来提高传输效率。本文针对传感器所感知的数据的波动幅度,采取有效的降熵变换,结合压缩编码,实现数据的无损高性能压缩。

2 相关工作

2.1 边缘计算

随着终端的智能化发展和大数据时代下呈几何增长的数据交互环境的日益完善,人们可以便捷地从终端获取日常资讯。图 1 表示基于双向数据流的边缘计算模型。作为数据的生产者和消费者的传感器、智能手机和其他物联网对象等可以向边缘侧发送请求,边缘侧接收请求且执行部分计算任务,再将计算结果反馈给终端,边缘侧一方面解决了终端功能多样性增强与终端能源不匹配的问题^[5];另一方面,边端的结合实现了高效、超低时延地处理工作。同时云端可以向边缘侧发送处理结果的调用请求及反馈结果,以借助云端的高计算资源完成更复杂的处理操作。总的来说,边缘计算就是贴近

终端或数据源头,在计算、存储等能力远小于云中心的情况下,提供一定最近端服务,快速响应需求。

边缘计算技术的应用领域非常广泛,涉及到方方面面的应用。比如,跨域数据共享存在的安全问题,即单个域内数据共享安全性有保障,而跨域数据共享安全风险高。因此 Fan 等^[6]利用边缘节点管理各个域,负责数据传输和处理,而边缘节点信任根在云端,从而实现高安全的跨域数据共享。Chen 等^[7]结合实际搭建的边缘平台来诠释边缘节点的功能,在设备层建立统一架构,实现信息交互的统一语义。网络域控制数据传输过程,数据域完成计算与存储,包括数据清洗和特征提取等操作,应用域将数据域分析结果加以应用。

总的来说,边缘计算的存在既缓解了云平台的压力,又为终端提供了更加高效便捷的服务。

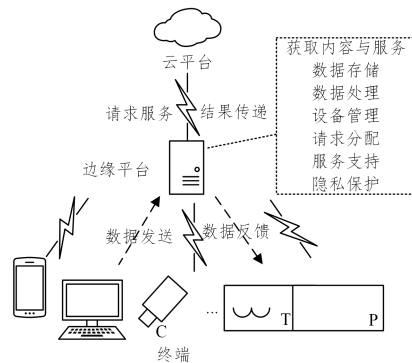


图 1 基于双向数据流的边缘计算模型

Fig. 1 Edge computing model based on bidirectional data flow

2.2 数据压缩

无论是为了存储还是传输,压缩的目的是为了减少数据量。关于在不同场景下对海量连续的实时数据的压缩处理研究目前也很多,Li 等^[8]提出一种贪心的哈尔小波概要数据结构高效压缩道路行驶数据,压缩率可达到 3%~7%,且能高效重构,重构后的每一个数据误差均不会超过设定的误差预值。Li 等^[9]提出一种基于分段自适应阈值的小波压缩算法来应对地铁交通拉门数据。该算法对数据进行自适应分段和自动调节各段阈值放大系数,兼顾了精度和大压缩比。Zhao 等^[10]为了更高效、安全地压缩智能电表测量数据,在基于边缘计算的配电网信息采集与监控架构下,利用经验模态分解算法对原始分布监测数据进行分解,并利用层次树中的集合划分算法对清洗后的数据进行压缩,但存在场景局限。在电力数据采集压缩的过程中,Lee 等^[11]为了有效压缩高频采集数据,提出一种基于频率选择的自动编码压缩模型,能在保证压缩比的条件下改善重建效果。而对于这种多元电力数据的压缩,Chowdhury 等^[12]提出自适应多元数据压缩算法,利用感知设备能够感知的不同变量之间的互相关性来降低数据的维数。随后,每个去相关流中的稀疏性被用于时间压缩。实验结果表明在实际通信中可以降低 98.5% 传输带宽需求。Azar 等^[13]提出一种小波变换提升的方案来压缩人体穿戴设备感知数据,该压缩方法对波动小的数据压缩效果很好,但对振动变化快的数据压缩效果很差。

当然,有时候考虑实现高精度重构会在很大程度上影响压缩性能,例如文献[14]中基于大型基础设施系统结构健康

监测数据,使用深度学习的方法提出一种数据压缩和重建的框架,虽然实现了高精度重构但压缩效果只有 10%。

而考虑执行算法的设备计算资源有限的问题时,文献[15]提出了邻域索引序列算法,通过利用相邻位的出现,动态地将较短长度的码字分配给输入序列中的每个字符,压缩比高达 89.13%,鲁棒性强。

在工业数据压缩方面,Park 等^[16]提出按照数据的趋势来处理工业数据,数据变化趋势具备更高价值。考虑到不同数据的不同变化趋势时,文献[17-18]分别在机泵在线监测和气候监测场景下,针对不同变化趋势的数据分别提出利用相似性匹配的改进旋转门算法和合理切换使用离散傅里叶变换基固定字典或 K-SVD 学习字典的数据压缩方案来减少该场景下的系统功耗。

而在物联网研究方面,往往重点考虑数据传输过程中的能耗,文献[19]中为减少数据传输能耗,在物联网设备中部署快速误差受限的有损压缩器,在数据处理中心重建数据,实验结果表明,在不影响数据质量和预测精度的情况下,传输的数据量减少了 103 倍。目前,考虑到工业数据采集时的噪声和误差对工业数据有损压缩的影响的研究有很多,无损压缩研究较少。而云端在用数据做训练处理时,需要使用原始数据,因此需保证原始数据可以被上传,无损压缩的方式可以得到很好的应用。无损压缩相比于有损压缩的计算冗余比更低,但能无失真重构数据。不同场景的数据不同,数据的性质决定压缩算法该具备的性能,目前研究的压缩算法在应用领域中都存在一定限制。

综上所述,在需要将工业现场的大量感知数据上传至云中心处理时,可以有效结合边缘计算技术,考虑边缘侧计算资源和存储资源较小的情况下,利用合适高效的无损压缩算法,来缩减需要传送到云端的数据大小,以节省数据传送带宽以及存储空间。

3 基于降熵变换的无损压缩方法

数据无损压缩的理论极限和数据的熵有关^[20]。无损压缩是利用数据的统计冗余进行压缩的,冗余大小与数据中每个符号的概率以及相关性有关。对于离散数据,数据呈独立等概率分布时,熵最大。熵的计算式如式(1)所示,其中, x 表示数据样本, i 为样本编号, n 是样本总数, $p(x_i)$ 为离散随机变量的概率质量函数。对于离散数据,数据呈独立等概率分布时,熵最大,如式(2)所示。感知数据间往往存在相关性,所以实际熵小于数据独立时计算的熵,实际的熵越小,数据间相关程度越高,数据可压缩程度越高。并且数据的概率分布越不均匀,熵就越小。所以使用一些数据变换的方式令原数据的概率分布变得不均匀即可减小数据熵,增加数据可压缩程度。将可降低原数据的熵的数据变换方式称为降熵变换。

$$H(\mathbf{X}) = -\sum_{i=1}^n p(x_i) \log_2 p(x_i) \quad (1)$$

$$H_{\max}(\mathbf{X}) = \log_2 n, p(x_1) = p(x_2) = \dots = p(x_n) = 1/n \quad (2)$$

数据压缩流程如图 2 所示,其中, σ_{DS} 表示输入数据差分序列的标准差,如式(3)所示; F 表示数据采集频率,可根据现场采集条件自定义。利用这两个参数将数据按照变化趋势划分开,再利用降熵变换方法,增加数据可压缩程度,实现边缘

侧数据的高性能压缩。

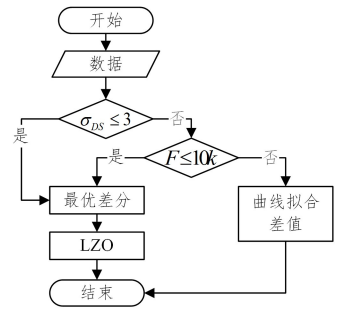


图 2 数据压缩流程

Fig. 2 Data compression process

感知数据间往往存在很高的相关性,这些相关性为我们提供了压缩的空间。为了更加高效地压缩数据,对工业现场感知数据进行分析,根据数据采集频率将其分为低频数据和高频数据。低频数据的数据波动趋势对数据压缩效率影响很大。所以,通过对数据的分析发现,感知数据大都服从或近似正态分布,某时刻感知数据值 d_t 与该时刻真正的数据值 r_t 之间存在微小的差异 ϵ_t ,可以表示为式(3),所以该时刻的 d_t 服从均值为 r_t ,方差为 ϵ_t 的正态分布。

$$d_t = r_t + \epsilon_t, \epsilon_t \sim N(0, \sigma_t^2) \quad (3)$$

所以,两个连续时刻的采样值 d_{t1}, d_{t2} 有如式(4)所示的关系。文献[21]中的数据分析表明,当感知数据满足 $\sigma_{r_{t1}-r_{t2}} \leq 3$,如式(5)所示,此时两个数据之间变化的差值近似于 0,数据波动幅度较缓,并且在一段时间内数据值保持不变。按照式(5),以数据差值序列标准差作为数据变化特征的第一划分条件,将低频数据中平缓波动的数据区分开来,再利用采集频率区别出高频振动数据。

$$\Delta d_t = (d_{t1} - d_{t2}) \sim N(r_{t1} - r_{t2}, \sigma_{t1}^2 + \sigma_{t2}^2) \quad (4)$$

$$\sigma_{DS} = \sqrt{D(\Delta d_t)} = \sqrt{E[(\Delta d_t - E(\Delta d_t))^2]} \leq 3 \quad (5)$$

图 2 中采取了两种数据降熵的方法。第一种是最优差分,该方法是基于差分变换的基础上实现的。差分变换是大多数变换方法中的重要步骤,它可以改变原数据的概率分布,增加数据冗余度,进而提高数据压缩效果。而最优差分的原理是:差分次数需要根据差分后的数据区间来定,分别计算第 $i-1$ 次差分和第 i 次差分的数据分布区间,比较两者的大小。若第 i 次差分后区间小于 $i-1$ 差分区间,则说明数据区间仍可收敛,可以继续下一次差分;反之,数据差分区间收敛至最小范围,之后再差分,数据区间将发散数据差分区间收敛以达最优差分次数,停止差分操作。每次差分处理前,将第一点的数据保存,方便数据重构。

但是,利用这种变换来提升数据压缩比存在局限性。在 $\sigma_{DS} > 3$ 且采集频率较低的数据中,存在当前时刻数据差值比原数据值大的情况,致使差分变换不能起到好的压缩效果。因此,我们采用另一种降熵方法,即阈值分段线性拟合差值,通过控制拟合数据与原数据间的差值在阈值内进行高效编码,来弥补最优差分在数据降熵方面的缺陷。感知数据可视为以时间 t 为自变量,以感知数据值 y 为因变量的函数。当使用一元线性回归模型进行拟合时,感知数据值可用式(6)表示:

$$y_t = \alpha_1 + \alpha_2 t + \epsilon_t, \epsilon_t \sim N(0, \sigma_t^2) \quad (6)$$

其中,用最小二乘法估计参数 α_1 和 α_2 并将其中有关 t 的求和按照数学公式转换,得到的 α_1' 和 α_2' 表示为式(7)和式(8)。其中, $S_1=\sum_{i=1}^n y_i, S_2=\sum_{i=1}^n n \times y_i$ 。最后拟合的数据值的计算式如式(9)所示:

$$\alpha_1'=\frac{S_1}{n}-\left(\frac{n+1}{2}\right)\alpha_2' \tag{7}$$

$$\alpha_2'=\frac{[12S_2-6(n+1)S_1]}{n(n^2-1)} \tag{8}$$

$$y_i'=\frac{6S_2-2(n+1)S_1}{n(n+1)} \tag{9}$$

拟合数据值 y_i' 可以直接由 S_1 和 S_2 推出,当已知某个时刻的 S_1 和 S_2 ,就可以推算出该时刻的拟合值。第一点和第二点数据由 S_1 和 S_2 得出,后面的数据与拟合值做差,作为传输数据。为了便于对拟合后差值进行编码,需要将拟合差值设定一个阈值。在设定阈值后,可以根据阈值设定的大小决定压缩编码的位数。

数据经过降熵变换后,再进行进一步压缩。将 LZO 算法作为数据降熵变换后的二级压缩方法,根据降熵变换后的数据压缩效果,决定是否用 LZO 做二次压缩。LZO 算法原理是判断当前字符串是否在已处理字符串中出现过,若出现过则为重复字符,计算重复长度(重复字符串的长度)和指回距离(先后两个相同字符串之间相隔的距离);若没出现过则为新字符。该压缩算法的原理流程如图 3 所示。

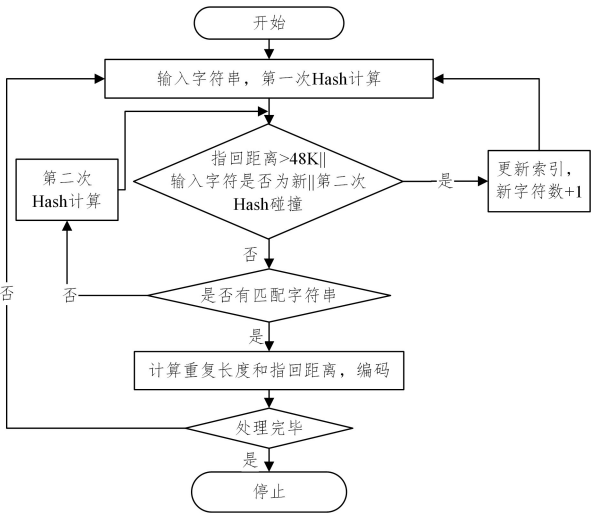


图 3 LZO 压缩流程图

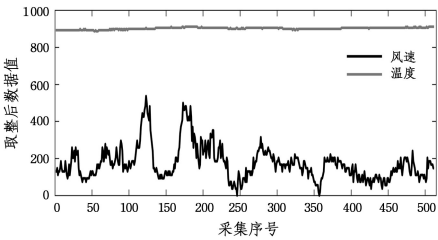
Fig. 3 LZO compression flow chart

4 实验验证

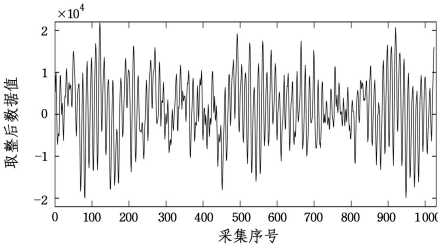
本文在 Linux 平台下采集数据,在 python 开发环境下测试压缩算法,在 Windows10 操作系统 Intel(R) core(TM) i7-10510U CPU @1.80GHz,运行内存为 12GB 的 PC 机上测试分析实验结果。

实验模拟数据的来源:1)2018 年 1 月在某工厂车间采集的环境温度数据,数据精确到 2 位小数;2)凯斯西储大学轴承数据中心(CWRU)¹⁾的单点驱动端和风扇端的正常基准数据,包含驱动端加速度计数据和风扇端加速度计数据,驱动端轴承实验以 12000 个样本/秒和 48000 个样本/秒的速度收集

数据,以每秒 12000 个样本的速度收集所有风扇端轴承数据,数据精确到 5 位小数;3)来自 EPFL 实验室提供的能源消耗数据集²⁾,数据精确到 1 位小数。图 5 为从以上 3 个数据集中选取的部分数据绘制的曲线图,数据已化为整数。



(a) 温度和风速数据分布曲线



(b) 驱动端加速度计的数据分布曲线

图 5 数据集中部分数据变化曲线

Fig. 5 Part of the data change curve of data set

将以上提到的公开数据集中的不同类型的数据,用不同的模拟终端通过不同的数据接入协议采集,利用这些数据做压缩实验。无损压缩算法常用的性能评价指标有压缩率、压缩时间、解压时间以及算法运行过程中占有内存大小。其中,压缩率的计算方式如式(10)所示:

$$\text{压缩率}=\frac{\text{原数据大小}-\text{压缩后数据大小}}{\text{原数据大小}} \tag{10}$$

将 LZO 与文献[10]中所提的压缩方法进行对比,实验分别测试了当数据量为 2048 和 4096 时的压缩率和压缩时间以及解压缩时间,性能测试结果如表 1 所列。

表 1 数据压缩率对比

Table 1 Data compression ratio comparison

算法	数据量	
	2048	4096
本文算法-温度	0.930	0.930
LZO-温度	0.857	0.857
文献[10]-温度	0.909	0.891
本文算法-加速度	0.822	0.822
LZO-加速度	0.601	0.613
文献[10]-加速度	0.482	0.497
本文算法-风速	0.770	0.770
LZO-风速	0.620	0.620
文献[10]-风速	0.765	0.752

从表 1 数据压缩率对比中,可以看出本文提出的算法在各种数据类型中压缩率均最高,并且随着数据量的增多,压缩效果仍然比较稳定,数据压缩率最高达到 93%,最低有 77%。相较于原 LZO,本文方法在压缩率上提升了 8%~22%。文献[10]中利用层次树中的集合划分算法对数据进行压缩,所以对于高频率变化的振动加速度计的数据压缩效果较差,压缩率不及 50%。

¹⁾ <https://csegroups.case.edu/bearingdatacenter/pages/download-data-file>

²⁾ <https://www.epfl.ch/labs/lisir/datasets/>

根据图 6、图 7 观察各算法的压缩时间和解压缩时间,因为 LZO 算法压缩和解压缩时间几乎可以不计,所以图中并未标出。对比本文算法和文献[10]中的压缩算法可知,本文算法的压缩和解压缩时间均低于文献[10]提出的算法,但差距不大。本文算法的压缩时间和解压缩时间均随着数据量的增多而加倍,而文献[10]中并未成倍增加,增加幅度较小。总体而言,本文算法在压缩/解压缩的速度上表现更好。

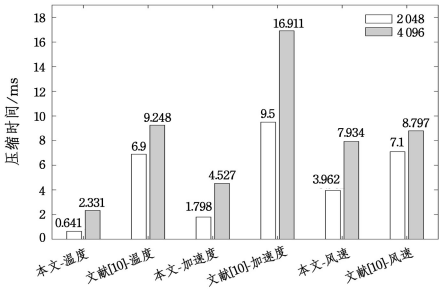


图 6 数据压缩时间对比

Fig. 6 Data compression time comparison

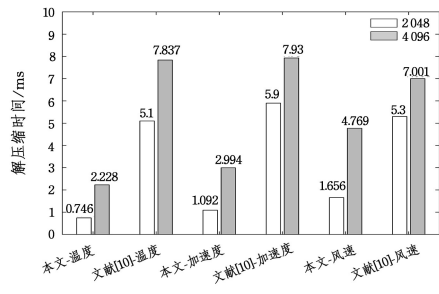


图 7 数据解压缩时间对比

Fig. 7 Data decompression time comparison

在进行以上压缩测试时,本文算法和文献[10]算法所占的运行内存结果如表 2 所列。针对边缘侧设备计算资源少的问题,用较少资源完成高效的任务非常重要。从表 2 中看出,本文算法在同等数据条件下,运行时占用内存非常少,相比于文献[10]的算法,本文算法计算资源少的设备适用性更高。

表 2 算法运行内存占用

Table 2 Algorithm running memory usage (单位: kB)

算法	数据量	
	2048	4096
本文算法-温度	48	76
文献[10]-温度	255	442
本文算法-加速度	264	528
文献[10]-加速度	689	1459
本文算法-风速	24	64
文献[10]-风速	277	608

使用 restful 的 POST 方式在同一局域网中上传数据,图 8 给出了上传不同文件大小的传输时间。通过 post 传输数据时,每增加 1 MB 文件进行传输,传输时间增加 1 s 左右,具体传输时间依据当前局域网网速确定。因此在该网络条件下,经过压缩后,数据传输量增加 82% 以上。最优差分法和曲线拟合差值均是可逆的变换,LZO 是无损压缩算法,所以解压缩过程就是压缩的逆过程,可无损恢复数据。实验对比了原数据和解压后的数据,误差为 0,说明该方法可以无损恢复数据。

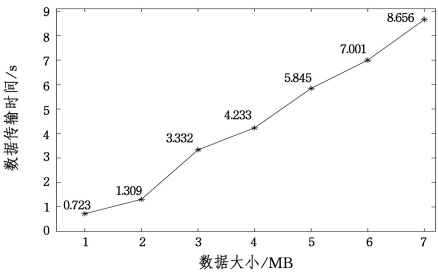


图 8 数据传输时间对比

Fig. 8 Data transmission time comparison

结束语 本文研究了在基于工业边缘计算场景下的工业感知数据压缩的方法。根据数据差值序列标准差判断数据的波动程度,借助采集频率将高频振动数据筛选出来,利用有效的降熵变换方法,即最优差分法和曲线拟合,改变原数据概率分布,增加数据可压缩程度。借助边缘设备模拟采集目前已有数据集的数据做算法压缩测试,并与 LZO 压缩算法进行对比,验证本文算法在压缩率方面的提升。当然,本文提出的方案也有些许不足,后续可以有针对性地进行改进,例如在曲线拟合上估计参数值过大的问题以及算法在编写上的优化。

智能制造推进“中国制造 2025”国家战略,而智能工厂也是智能制造的重要实践领域。随着智能化的推进,我们能够获取的数据量会越来越大。本研究为工厂的数据存储或通信传输提供了很好的数据压缩方法。

参考文献

[1] ZHANG J L,ZHAO Y C,CHEN B,et al. Survey of Edge Computing Data Security and Privacy Protection Research[J]. Journal on Communications,2018,39(3):1-21.

[2] ZHAN X,GUO H,HE X Y,et al. Research on Security Risk Assessment Method of State Grid Edge Computing Information System [J]. Computer Science,2019,46(S2):428-432.

[3] Cisco cloud index supplement. Cloud readiness regional details white paper [R/OL]. <https://blogs.cisco.com/tag/global-cloud-index>.

[4] XU D,LI Q,ZHU H. Energy-Saving Computation Offloading by Joint Data Compression and Resource Allocation for Mobile-Edge Computing [J]. IEEE Communications Letters, 2019, 23(4):704-707.

[5] WEI X, MVR; An Architecture for Computation Offloading in Mobile Edge Computing[C]// 2017 IEEE International Conference on Edge Computing (EDGE). Honolulu, HI, 2017: 232-235.

[6] FAN K,PAN Q,WANG J,et al. Cross-Domain Based Data Sharing Scheme in Cooperative Edge Computing[C]// 2018 IEEE International Conference on Edge Computing (EDGE), San Francisco,CA,2018:87-92.

[7] CHEN B, WAN J,CELESTI A, et al. Edge Computing inIoT-Based Manufacturing [J]. IEEE Communications Magazine, 2018,56(9):103-109.

[8] LI S Q,YUE Y,LI Q Q, et al. A streaming data compression method based on Haar wavelet synopses[J]. Geomatics and Information Science of Wuhan University, 2021, 46(8): 1216-1223.

- [9] LI C X, JIANG Q L, LU F, et al. Data compression method of subway sliding door based on segmented adaptive threshold wavelet[J]. Journal of Harbin Institute of Technology, 2020, 52(9):101-106.
- [10] ZHAO Z, LU J, DONG D, et al. Intelligent Terminal Data Compression Method Based on Edge Computing[C]//Proceedings of the 3rd International Conference on Information Technologies and Electrical Engineering, 2020:579-584.
- [11] LEE J, YOON S, HWANG E. Frequency Selective Auto-Encoder for Smart Meter Data Compression[J]. Sensors, 2021, 21(4):1521.
- [12] CHOWDHURY M R, TRIPATHI S, DE S. Adaptive multivariate data compression in smart metering Internet of Things[J]. IEEE Transactions on Industrial Informatics, 2020, 17(2):1287-1297.
- [13] AZAR J, DARAZI R, HABIB C, et al. Using dwf lifting scheme for lossless data compression in wireless body sensor networks [C]//2018 14th International Wireless Communications & Mobile Computing Conference(IWCMC). IEEE, 2018:1465-1470.
- [14] NI F T, ZHANG J, NOORI M N. Deep learning for data anomaly detection and data compression of a long-span suspension bridge[J]. Computer-Aided Civil and Infrastructure Engineering, 2020, 35(7):685-700.
- [15] UTHAYAKUMAR J, VENGATTARAMAN T, DHAVACHELVAN P. A new lossless neighborhood indexing sequence (NIS) algorithm for data compression in wireless sensor networks[J]. Ad Hoc Networks, 2019, 83(FEB.):149-157.
- [16] PARK J, PARK H, CHOI Y. Data compression and prediction using machine learning for industrial IoT[C]//2018 International Conference on Information Networking (ICOIN). Chiang Mai, 2018:818-820.
- [17] YAO X Z, SHANG J F, CAO J J, et al. A data compression method based on similarity matching for machine and pump monitoring analog data[J]. Modular Machine Tool and Automatic Manufacturing Technology, 2020(8):83-87, 91.
- [18] ZHENG Y L, ZHAO Y, ZHAO Y D, et al. Compressed sensing method of forest microclimate monitoring data based on switching dictionary[J]. Transactions of the Chinese Society of Agricultural Machinery, 2019, 50(11):193-199.
- [19] AZAR J, MAKHOUL A, BARHAMGI M, et al. An energy efficient IoT data compression approach for edge machine learning [J]. Future Generation Computer Systems, 2019, 96 (JUL.): 168-175.
- [20] FU Z Y. Information Theory-- basic Theory and Application (Second Edition) [M]. Beijing: Electronic Industry Press, 2007.
- [21] REN X J. Research on lossless compression algorithm of sensing data in wireless sensor networks based on entropy drop transform [D]. Illinois: Northwestern University, 2011.



ZHANG Xiao-mei, born in 1977, Ph.D., professor, postgraduate supervisor. Her main research interests include intelligent information processing and information security



LI Da, born in 1980, Ph.D., associate professor. His main research interests include statistical signal processing, computer vision, machine learning, and pattern recognition.