

# 面向无线传感网络应用的改进 LZW 算法



倪晓军 余成豪

南京邮电大学计算机学院 南京 210046

**摘要** 在无线传感网络通信中,传感器数据需要通过无线设备发送给上位机。随着终端传感器传输数据量的增大,无线设备的发送能耗逐渐加大。在不利于及时维护的复杂环境中,这将导致无线通讯设备过早失效从而使得通讯中断。因此需要先将传感器采集到的数据进行压缩,减小发送数据量。在分析传感器数据特点和传统的 LZW(Lempel-Ziv-Welch)压缩算法的基础上,提出了一种面向无线传感网络应用的改进 LZW 算法。该算法首先对采集到的传感器相邻数据进行差值预处理,以提高数据项的重复率;然后选择大小合适的字典,在字典上用哈希存储的方式代替传统的顺序存储,以改进字典更新方式,当检测到压缩率降低时更新字典,并保存常用单字符,释放字典空间,达到数据压缩的目的。实验数据显示,与传统的 LZW 算法相比,改进的 LZW 算法使得有序传感器数据的压缩率最高降低 40%,减小了所需发送数据的数据量,压缩速度也提高了近 10 倍,证明了面向无线传感网络应用的改进 LZW 算法是有效可行的。

**关键词**: 无线传感网络; 压缩算法; LZW 算法; 数据预处理; 压缩率

**中图分类号** TP301.6

## Improvement of LZW Algorithms for Wireless Sensor Networks

NI Xiao-jun and SHE Xu-hao

School of Computer Science, Nanjing University of Posts and Telecommunications, Nanjing 210046, China

**Abstract** In wireless sensor network communication, sensor data need to be sent to the host computer through the wireless device. With the increase of the amount of data needed to be transmitted by the terminal sensors, the energy consumption of wireless devices is gradually increasing. A complex environment that is not convenient for timely maintenance lead to premature failure of wireless communication equipment and communication interruption. Therefore, it is necessary to compress the data collected by the sensor to reduce the amount of data sent. Based on the analysis of sensor data characteristics and traditional LZW (Lempel-Ziv-Welch) compression algorithm, an improved LZW algorithm for wireless sensor network applications is proposed. Firstly, the algorithm preprocesses the adjacent data collected from sensor to calculate the difference, so as to improve the repetition rate of the data items. Then, the appropriate dictionary size is selected and the traditional order memory is replaced with the hash memory in the dictionary, so as to improve the way of dictionary updating. When the compression rate is decreased, the proposed algorithm updates the dictionary, and saves the common single character to release the dictionary space, for data compression. The experimental results show that compared with traditional LZW algorithm, the improved LZW algorithm reduces the compression rate of the ordered sensor data by up to 40%, and reduces the amount of data needed to send. The compression speed is also increased by nearly ten times, which proves that the improved LZW algorithm for wireless sensor network applications is effective and feasible.

**Keywords** Wireless sensor network, Compression algorithm, LZW algorithm, Data preprocessing, Compression ratio

### 1 引言

无线传感网络(Wireless Sensor Network)是一种分布式的,在末梢由大量传感器节点组成的无线自组织网络系统<sup>[1]</sup>。在无线传感网络通信中,传感器数据首先从网络末端的终端设备发送到负责选路的路由器设备,再通过路由器设备发送到协调器设备汇总,上位机通过协调器取得数据。在此期间,数据的成功传输依靠无线传输设备的正常工作,一般情况下传输数据量越大,无线设备所消耗的能量越多<sup>[2]</sup>。无线传输

设备节点在其能量耗尽时会严重阻碍数据的传输,因此研究 WSN 发送数据的精简方法显得尤其必要。

LZW 压缩(LZW compression)由 Abraham 等发明,其基于压缩过程中同步生成的字典把文件压缩成较小的文件,是一种非常适合大量重复数据的无损压缩算法。该算法将重复字符存储进同步生成的字典中并赋予标号,当再次遇到这组重复字符时只需输出字典的标号即可。其逻辑简单,只需进行字典的读写和查找操作就可以完成数据的压缩<sup>[3]</sup>。但 WSN 中终端节点一般资源比较匮乏且处理能力有限,致使字



00,00,01...)。在处理后的待压缩项中会出现长串的重复项,降低了数据的变化率,极大提高了数据的重复率,使处理后的数据更加适合于LZW算法。

### 3.2.2 字典存储策略

在LZW压缩算法中,输出的压缩编码实际上是数据中的重复字符串在字典里的地址<sup>[14]</sup>,所以将数据经过LZW压缩算法压缩后,输出的数据大小与算法中自动生成的字典大小有关<sup>[15]</sup>。当设置的字典容量较大时,查找命中率提升,压缩率也会提升。但字典过大时,查找字符串的耗时较长,会造成压缩速度过慢;当设置的字典较小时,字典的匹配率不高,会影响后续的压缩效果<sup>[16]</sup>。

经过实验对比,传统LZW算法的字典大小与压缩比和压缩时间的对应关系如图2所示。

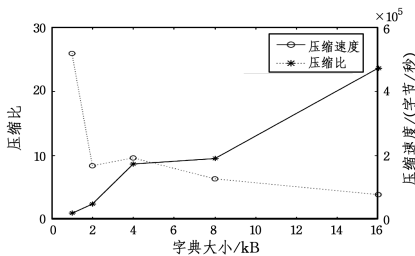


图2 LZW算法的字典大小与压缩比和压缩时间的对应关系

Fig. 2 Relationship among dictionary size of LZW algorithm, compression rate and compression time

从图2可以看出,字典大小选择4K时,压缩比和压缩速度最佳。但算法不论选择多大的字典,大量涌入的数据很快就会将字典填满,所以字典的更新方式是关键。传统LZW算法在字典被填满后不对字典进行实时更新,必然会影响后续数据的压缩操作<sup>[17]</sup>。然而,经过差值处理的传感器数据重复率较大,频繁地更新字典会让字典中的高频字符被丢弃,使得压缩率降低。对此,结合传感器采集的数据的特点,采用改进的字典更新方式:在压缩操作中,当字典中查不到待匹配字符串时,将当前读取的单字符在字典中存储位置的标志位VALUE设置为1。当字典被填满后将表示字典已经被填满的标志位FILL置1。在字典被填满后不立即更新字典,每读取指定字节数的数据就比较一次压缩率,当检测到压缩率升高时,再进行字典的更新操作。字典标志位设置规则的伪代码如下所示。

#### 算法1 标志位设置算法

IF P+C is in the dictionary THEN

P ← P+C

ELSE

IF the dictionary is not full THEN

output the code for P

add P+C to the dictionary

set the sign VALUE of character C in the dictionary as 1

P ← C

ELSE

full sign FILL ← TRUE

output the code for P

P ← C

END of IF

END of IF

更新字典时,只保存做过标记的单字符,将剩余空间释放并加入压缩处理的字符串存储空间中。更新字典部分的伪代码如算法2所示。

#### 算法2 字典更新算法

IF the first dictionary initialization THEN

initialize all the single character contents

empty the rest contents

ELSE

only initialize the single character contents which VALUE = 1

empty the rest contents

END of IF

解压时不需要考虑更新字典的时间,只需要根据压缩算法中的更新字符来完成更新操作,更新字典的步骤与压缩算法相同。在写入字典时将字符串的第一个字符在字典中所存储的位置标记置1即可。

该改进算法只在写入字典时设置标志位,在初始化时根据标志位初始化单字符,并在程序运行时检测输入、输出数据的个数,并未增加算法的复杂度。

字典采用哈希散列表的方式进行存储。传统的字典读写方式为顺序读写,时间复杂度为 $O(n^2)$ ,引入哈希表能有效地提高字典的查找效率以及整体的压缩速度<sup>[18]</sup>。哈希表的每一表项存放与该表项哈希值相同的数据,其复杂度为 $O(1)$ 。经实验,当字典大小为4K时,采用哈希表存储的字典算法的运行时间只有传统LZW算法的25%~50%,使用哈希表作为字典的存储方式极大地减少了程序运行时间,符合无线网络下数据处理的环境要求。

表2 LZW算法采用传统字典和改进字典的压缩效果比较

Table 2 Comparison of compression effect between LZW traditional and improved dictionary in LZW algorithm

字典	待压缩/字节	压缩后/字节	压缩率/%	压缩速度/s
传统字典	1130935	484791	42.27	3.55
改进的字典	1130935	74043	6.55	0.57
传统字典	540574	62499	11.56	2.34
改进的字典	540574	28077	5.19	0.26

由表2可以看出,采用改进的字典存储方法后,压缩率大幅降低,压缩速度大幅提高。

### 3.2.3 整体改进思路

改进后的LZW算法维护4个变量P,C,C<sub>cur</sub>和C<sub>pre</sub>,P同传统LZW算法一样表示未被编码的字符串项,C<sub>cur</sub>表示当前读入的单字符数据,C<sub>pre</sub>表示上一次读入的单字符数据,C表示待压缩字符。改进后的LZW算法的伪代码如算法3所示。

#### 算法3 改进后的LZW算法

Initialize the dictionary

P ← 0

C ← 0

C<sub>cur</sub> ← 0

C<sub>pre</sub> ← 0

WHILE dictionary doesn't reach the end THEN

IF C<sub>cur</sub> = 0 THEN



而提高字典的利用率,降低压缩率。

改进的 LZW 算法对传感器数据具有很强的适应性,能根据数据间的规律性地减少无线传感网中需要传输的数据。该算法也要求传感器数据的误差尽量低,即采集数据时噪点应尽量少,以免破坏传感器数据的规律性,使压缩率相对增大。对于前后关联性不大、重复率不高或冗余度不高的数据,数据预处理不能有效地增加数据的重复性,使得改进的压缩算法的压缩效果不明显,这是值得进一步研究的地方。

## 参 考 文 献

- [1] CHUNAWALE A, SIRSIKAR S. Minimization of average energy consumption to prolong lifetime of Wireless Sensor Network [C]//2014 IEEE Global Conference on Wireless Computing & Networking (GCWCN). Lonavala, 2014: 244-248.
- [2] IRMAK E, KÖSE A, GÖÇMEN G. Simulation and ZigBee based wireless monitoring of the amount of consumed energy at smart homes[C]//2016 IEEE International Conference on Renewable Energy Research and Applications (ICRERA). Birmingham, 2016: 1019-1023.
- [3] KUMAR R, MALIK A, SINGH S, et al. Reversible data hiding scheme for LZW codes using even-odd embedding strategy[C]//2016 International Conference on Computing, Communication and Automation (ICCCA). Noida, 2016: 1399-1403.
- [4] MARKSTEINER S, JIMENEZ V J E, VALIANT H, et al. An overview of wireless IoT protocol security in the smart home domain[C]//2017 Internet of Things Business Models, Users, and Networks. Copenhagen, 2017: 1-8.
- [5] LI S, LIAN G, PENG Z. Research on the loss-less Compression algorithm of the ultrasonic testing of rail[C]//2015 Symposium on Piezoelectricity, Acoustic Waves, and Device Applications (SPAWDA). Jinan, 2015: 458-461.
- [6] MANJULA Y, SHIVAKUMAR K B. Enhanced secure image steganography using double encryption algorithms[C]//2016 3rd International Conference on Computing for Sustainable Global Development (INDIACom). New Delhi, 2016: 705-708.
- [7] SANGEETHA M, BETTY P, KUMAR G S N. A biometric iris image compression using LZW and hybrid LZW coding algorithm[C]//2017 International Conference on Innovations in Information, Embedded and Communication Systems (ICIIECS). Coimbatore, 2017: 1-6.
- [8] KAUR G, KAD S. Improved Bee-Inspired routing protocol using Lzw based lossless compression[C]//2015 2nd International Conference on Recent Advances in Engineering & Computational Sciences (RAECS). Chandigarh, 2015: 1-6.
- [9] PAN G, HE J, WU Q, et al. Automatic stabilization of Zigbee network[C]//2018 International Conference on Artificial Intelligence and Big Data (ICAIBD). Chengdu, 2018: 224-227.
- [10] ALAM M A, FAKHRUL A, ATEEQ-UR-RAHMAN S, et al. Faster Image Compression Technique Based on LZW Algorithm Using GPU Parallel Processing[C]//2018 Joint 7th International Conference on Informatics, Electronics & Vision (ICIEV) and 2018 2nd International Conference on Imaging, Vision & Pattern Recognition (icIVPR). Kitakyushu, Japan, 2018: 272-275.
- [11] PRABU S, MAHESWARI R. Improving wireless sensor network lifespan by efficient routing algorithm and LZW compression[C]//International Conference on Information Communication and Embedded Systems (ICICES2014). Chennai, 2014: 1-5.
- [12] MALIK A, KUMAR R, SINGH S. Reversible Data Hiding Scheme for LZW Codes using LSB Flipping Strategy[C]//Proceedings of the International Conference on Advances in Information Communication Technology & Computing (AICTC'16). New York: ACM, 2016.
- [13] NAGPURKAR A W, JAISWAL S K. An overview of WSN and RFID network integration[C]//2015 2nd International Conference on Electronics and Communication Systems (ICECS). Coimbatore, 2015: 497-502.
- [14] LI Y M, LIANG Y. Temporal Lossless and Lossy Compression in Wireless Sensor Networks[J]. ACM Transactions on Sensor Networks, 2016, 37: 1-35.
- [15] KUMAR R, CHAND S, SINGH S. An optimal high capacity reversible data hiding scheme using move to front coding for LZW codes[J]. Multimedia Tools and Applications, 2019, 78(16): 22977-23001.
- [16] LI S H, YEN D C, CHUANG Y P. A Real-Time Audit Mechanism Based on the Compression Technique[J]. ACM Trans. Manage. Inf. Syst., 2016, 4: 1-25.
- [17] SINGH S, PANDEY P. Enhanced LZW technique for medical image compression[C]//2016 3rd International Conference on Computing for Sustainable Global Development (INDIACom). New Delhi, 2016: 1080-1084.
- [18] GALHOTRA K, KAUR K. Extending IAMCTD using data fusion and lossless data compression for UWSN[C]//2015 International Conference on Futuristic Trends on Computational Analysis and Knowledge Management (ABLAZE). Noida, 2015: 526-531.



**NI Xiao-jun**, born in 1969, master, associate professor. His main research interests include design and implementation of embedded system and its application in communication field and wireless sensor network field.