

# 地震交换标准数据的压缩技术和数据解压\*

杨周胜, 杨晶琼, 姚 远

(云南省地震局, 云南 昆明 650224)

**摘要:** 研究地震数据存储及交换的 SEED 和 miniSEED 格式标准, 对存储格式中的 Steim1 和 Steim2 的压缩方法和编码方式进行了深入分析, 对比了国内外 SEED 格式的差异, 利用 MATLAB 软件实现了 SEED 和 miniSEED 格式压缩数据的解压和应用。

**关键词:** SEED; miniSEED; Steim1; Steim2; 数据压缩; 编码方式; 数据解压

**中图分类号:** P315.63

**文献标识码:** A

**文章编号:** 1000-0666(2019)01-0144-06

## 0 引言

地震数据交换标准 (Standard for the Exchange of Earthquake Data, 简称 SEED), 是国际数字地震数据交换的一种标准格式, miniSEED 格式 (王晓磊等, 2016) 是 SEED 格式的简化版。两种格式数据在测震观测中被广泛使用, miniSEED 格式主要用于单台, 形成台站卷, 而 SEED 格式主要用于台网和台阵, 形成台网卷。早期的中国数字地震台网 (China Digital Seismograph Network, 简称 CDSN 台) 的地震记录就开始使用 SEED 格式 (和跃时, 2007), 我国的测震观测数据从“十五”建设开始即与国际接轨, 地震数据存储和交换采用了 SEED 格式。SEED 格式将地震波形数据进行了压缩, 有效地解决了存储效率, 但同时也带来了应用方面的不便。一般使用者通过两种方式来实现应用: 一是利用公开的转换软件, 如 rdSEED (何加勇, 2009), 进行一次转换成 ASCII 码, 然后再编写软件转换成自己需要的格式; 二是研究压缩技术编写软件, 直接转换为设定的格式。第一种方式繁杂, 使用起来不方便, 灵活性不强; 第二种方式, 由于标准和文献的描述和地震波形数据的实际存在一定的出入, 实现起来有非常大的困难。检索以往文献, 多侧重于标准、压缩技术等方面的探讨, 对具体的压缩方法和解压转换方面的论述寥

寥无几。本文基于第二种方式, 深入研究这两种格式数据, 找出标准和文献描述与实际地震记录的差异, 利用 MATLAB 实现了压缩数据的解压和各种转换。

## 1 SEED 和 miniSEED 格式

### 1.1 格式体

SEED 格式体包含控制头段 (ASCII 格式) 和时间序列 (二进制), 其中, 控制头段包含关于卷、台站、通道和数据的辅助信息, 时间序列 (二进制) 包含原始数据以及嵌入的与特定通道和时间有关的辅助信息。

### 1.2 控制头段

控制头段共 4 种: 卷索引控制头段, 包含关于数据的时间、逻辑记录长度、该逻辑卷的格式版本以及对台站控制头段和时间片控制头段的索引信息 (朱文林等, 1994a, b); 缩略语字典控制头段, 包含用在其它控制头段中的缩略语的定义; 台站控制头段, 提供有关台站和它的所有通道的信息; 时间片控制头段, 标识随后的时间序列所处的时间段。

### 1.3 数据记录

数据记录紧接着控制头段, 用于存储数据的部分。每个记录长度为 512 字节, 若干个记录组成一个数据记录。miniSEED 格式记录中, 3 个方向的数

\* 收稿日期: 2018-10-15.

基金项目: 地震科技星火计划项目“新震级国家标准在云南台网的应用 (XH18043)”资助.

据一起存储，不同方向的记录交替存放，某一方向可连续使用几个记录，数据结构如图 1 所示。

一帧 64 字节存储固定头段	512 字节 一个记录	某一台站 某一方向 的记录
64 字节存储数据		
64 字节		
64 字节		
64 字节		
64 字节		
64 字节		
64 字节		

图 1 单个 512 字节数据记录的数据结构  
Fig. 1 Data structure of a single 512-byte data record

每个记录中又分为 8 帧，每帧 64 字节。记录的第一帧，前 48 个字节称为固定头段区，统一存放固定头段信息，其中，前 20 个字节采用 ASCII，后面的 28 个字节采用二进制存储。固定头段区从数据记录的第一个字节开始，前 6 个字节为序号，第七个字节是一个 ASCII 字符“D”，表示其为一数据记录，第八个字节为 ASCII 字符空格。接下来的 12 个字节包含该记录的台站、位置及通道标识符信息。头段的二进制部分的内容为：记录起始时间、样本数目、采样率因子、采样率乘数、后面的子块数目、数据开始偏移量、第一个子块偏移量等信息。

第一帧中固定头段区后的 16 个字节，采用二进制存储，用于存储子块，一般可存放 2 个子块，如果子块数目较多，可采用可变头段（万永革等，1995）的方式进行存放。固定头段的后 16 个字节根据实际进行子块存放，如有 2 个子块，则 16 个字节全部使用；如果只有 1 个子块，则其余字节为空闲；如果不接任何子块，则后续 16 个字节均为空闲。固定头段的内容如表 1 所示。

SEED 格式记录是对 miniSEED 记录的重新整理，3 个方向独立，并以 4 096 字节为一记录，实际记录中是将同一方向连续 8 个 512 字节的记录串起来，形成一个 SEED 记录，数据结构见图 2。

2 Steim 压缩算法

Steim 压缩算法是地震数据交换标准的核心压缩技术，其方法是在一个记录中，先存放 1 个 32

表 1 固定头段区信息  
Tab. 1 Fixed head section information

序号	字段名	类型	长度（字节）
1	数据记录序号	ASCII	6
2	数据头段/质量指示符	ASCII	1
3	保留字节（“△”）	ASCII	1
4	台站标识符	ASCII	5
5	位置标识符	ASCII	2
6	通道标识符	ASCII	3
7	台网编码	ASCII	2
8	记录起始时间	BIN	10
9	样本数目	BIN	2
10	采样率因子	BIN	2
11	采样率乘数	BIN	2
12	活动标志	BIN	1
13	输入/输出和时钟标志	BIN	1
14	数据质量标志	BIN	1
15	后面的子块数目	BIN	1
16	时间校正值	BIN	4
17	数据开始偏移量	BIN	2
18	第一个子块偏移量	BIN	2

512 字节	首个SEED记录 (4 096字节)	某一台站某一 方向存储对应 的全部记录
512		
512		
512		
512		
512		
512		
512		
...	...	
512	最后一个SEED 记录 (4 096字节)	
512		
512		
512		
512		
512		
512		
512		

图 2 SEED 记录格式数据结构  
Fig. 2 Data structure of SEED record format

位数值的前向积分常数和 1 个 32 位数值反向积分常数, 后接连续记录的差值, 由于差值较小, 占用的数据位也相应变少, 这样可以显著地压缩数据, 节省存储空间。早期采用 Steim1 压缩算法 (李铂等, 2009), 由于方法的改进, 后来出现了压缩效率更高的 Steim2 算法 (王洪体等, 2004)。

### 2.1 Steim1 压缩算法

地震记录的两个连续样本之间的差值一般比较小, 只有在大地震时, 其差值才可能大于 8 位二进制数所能表示的最大值, 对这种极少发生的情况, 可以使用一个 2 字节或 4 字节数值来表示样本差值。Steim1 压缩算法就是一种编码方式, 它确定样本差值采用的字节数, 对于 99% 以上的地震记录, 样本差值只用 1 个字节存储。因此, Steim1 压缩方式节约了存储, 压缩比可达 3.5:1。Steim1 压缩方式的存储格式为: 前向积分常数、反向积分常数、两者之间的全部差值, 反向积分常数可对差值进行校核。

512 字节的数据记录中, 第一帧存放固定头段信息, 后 7 帧存放地震波形数据, 每帧均为 64 字节。存放数据的每一帧的前 4 个字节, 存有 16 个 2 位广义字节编码  $C_i$  ( $i$  取值 0~15), 每个  $C_i$  对应 4 个字节,  $C_i$  值确定对应 4 个字节的存储的差值数量。

$C_i = 00$  (二进制): 对应的 4 个字节存放特殊值, 如非数据信息、广义字节编码、前向积分常数、反向积分常数。

$C_i = 01$  (二进制): 对应的 4 个字节存放 4 个 1 字节差值。

$C_i = 10$  (二进制): 对应的 4 个字节存放 2 个 2 字节差值。

$C_i = 11$  (二进制): 对应的 4 个字节存放 1 个 4 字节差值。

$C_0$  对应于每一帧的前 4 个字节, 总是存放编码位, 其值等于 00 (二进制)。紧接固定头段后的第一个数据帧中,  $C_0 = C_1 = C_2 = 00$  (二进制);  $C_1$  对应的 4 个字节存放前向积分常数,  $C_2$  对应的 4 个字节存放反向积分常数。

地震波形记录的 Steim1 压缩数据格式见图 3,  $W_0$  对应广义字节编码,  $X_0$  对应前向积分常数,  $X_n$  对应反向积分常数,  $D_0 \sim D_3$  为样本差值。



图 3 Steim1 地震数据压缩格式图

Fig. 3 Compression format diagram of the Steim1 seismic data

### 2.2 Steim2 压缩算法

随着对地震波形数据研究的深入, 发现连续样本差值可以用小于 1 个字节来表示, 最小的差值可以用 4 个字节表示, 出现了 Steim2 算法。

Steim2 算法是将每个  $C_i$  对应的 4 个字节 (32 位) 的差值存储进行改进, 取高端 2 字节用于压缩算法的进一步解码。这 2 个字节称为 “dnib”, 即 “细解码” 广义字节。用后 30 个字节存储样本差值, 进一步提高了存储效率。

$C_i = 00$  (二进制): 同 Steim1, 存放特殊值, 如非数据信息、广义字节编码、前向积分常数、反向积分常数。

$C_i = 01$  (二进制): 同 Steim1, 存放 4 个 1 字节差值。

$C_i = 10$  (二进制): 依据 “细解码” dnib 的值确定存储差值的个数和位数。

dnib = 01 (二进制): 存放 1 个 30 字节样本差值。

dnib = 10 (二进制): 存放 2 个 15 字节样本差值。

dnib = 11 (二进制): 存放 3 个 10 字位样本差值。

$C_i = 11$  (二进制): 依据“细解码” dnib 的值确定存储差值的个数和位数。

dnib = 00 (二进制): 存放 5 个 6 字位样本差值。

dnib = 01 (二进制): 存放 6 个 5 字位样本差值。

dnib = 10 (二进制): 存放 7 个 4 字位样本差值。

如果连续 7 个差值都介于  $-8 \sim 7$ , 那么就将 7 个差值放在一个 4 字节中, 使用 32 个字位中的 28 个字位。样本差值为  $-16 \sim -8$  或  $7 \sim 15$  时用 5 个字位表示; 样本差值为  $-32 \sim -16$  或  $15 \sim 31$  时用

6 个字位表示; 样本差值在  $-128 \sim -32$  或  $31 \sim 127$  时, 用 8 个字位 (即 1 个字节) 表示; 样本差值在  $-512 \sim -128$  或  $127 \sim 511$  时, 用 10 个字位表示; 样本差值在  $-16\,384 \sim -512$  或  $511 \sim 16\,383$  时, 用 15 个字位表示; 样本差值在  $-2^{29} \sim -2^{14}$  或  $2^{14} - 1 \sim 2^{29} - 1$  时, 用 30 个字位表示。

地震波形记录的 Steim2 压缩数据格式见图 4,  $W_0$  对应广义字节编码,  $X_0$  对应前向积分常数,  $X_n$  对应反向积分常数,  $D_0 \sim D_6$  为样本差值。

3 压缩数据的解压

SEED 格式存储的数据有效地提高了存储效率, 节省了存储空间。但给一般的使用者带来许多不便, 多数是利用转换软件来拓展使用, 应用受到一定的限制。本文着重介绍如何针对压缩数据 (姚立平, 1993) 进行数据读取、转换和应用。

3.1 数据存储字序

计算机存储数据时, 存在不同的字节序。通常的字节序有 LE (Little endian) 和 BE (Big endian) 两种。LE 字节序是将低序字节存储在起始地址; BE 字节序是将高序字节存储在起始地址。每个字节中的字位存储顺序, 即字位序只有固定一种方式, 类似 BE 字节序, 即高序字位存储在起始地址。SEED 格式的数据存储采用 LE 字节序, 具体见图 5。常用软件系统的读取命令, 均采用从低地址到高地址的读取顺序, 见图 6。数据解压时, 字位存储和读取的顺序相反, 是数据解压的关键所在。



图 4 Steim2 地震数据压缩格式图  
Fig. 4 Compression format diagram  
of the Steim2 seismic data

4 字节二进制存储字节序和字位序				
字节存储顺序	第一字节	第二字节	第三字节	第四字节
字位存储顺序	76543210	76543210	76543210	76543210

图 5 SEED 格式中存储的字节序和字位序  
Fig. 5 Endian and word order stored in SEED format

一般命令读出字节序和字位序				
字节读取顺序	第一字节	第二字节	第三字节	第四字节
字位读取顺序	01234567	01234567	01234567	01234567

图 6 软件读取命令的字节序和字位序  
Fig. 6 Endian and word order of software  
reading commands

### 3.2 数据读取方法

在了解地震压缩数据的压缩方式和字节存储方式后,可利用能够读取二进制的系统编制软件进行数据读取。完全可以直接读取和应用数据,如进行数据处理、调用、转换、显示等。读取软件编制时,重点是循环的设置,对于 SEED 格式需设置 5 层循环嵌套,涉及台站方向、记录编号、miniSEED 记录编号、数据帧、帧内字节;对于 miniSEED 格式,则仅需 2 层循环嵌套即可。对于 SEED 格式的循环嵌套见图 7。

本文利用 MATLAB 系统,编制了数据读取软件。软件可直接读取地震压缩数据,免去转换,直接进行各种应用,同时也可转换为其它数据格式。图 8 即是软件读取 SEED 数据并绘制的地震记录图。

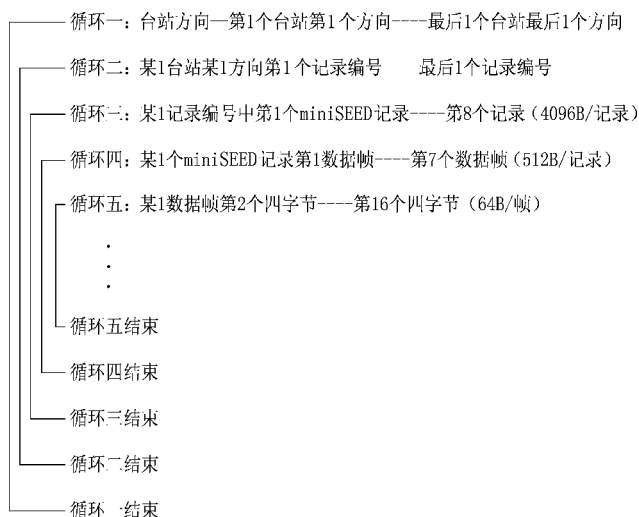


图 7 SEED 格式数据读取循环框图

Fig. 7 Cycle diagram of SEED format data reading

## 4 结论

本文通过对地震交换数据标准格式的研究,获得如下认识:

(1) 国内外 SEED 格式数据存在差异,台站卷的头文件一致,台网卷在头文件中存在差异,国内的头文件不含震源信息子块。部分国外仪器形成的 miniSEED 文件中,压缩格式采用 steim1 算法。

(2) 地震记录数据量巨大,选择压缩格式非

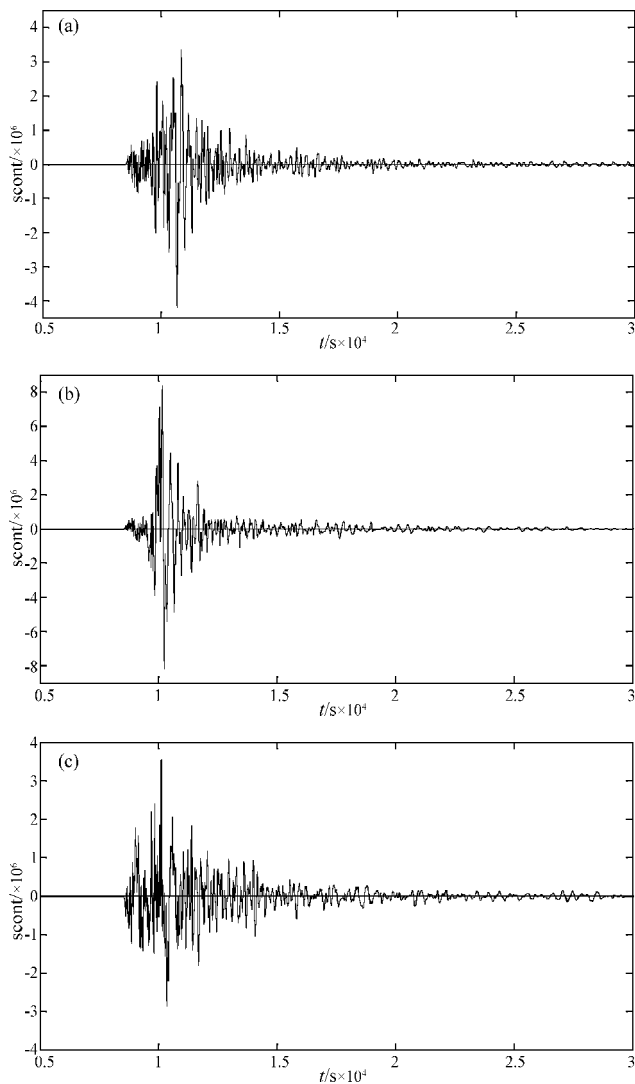


图 8 景谷台记录的墨江 5.9 级地震 UD 向 (a)、NS 向 (b) 和 EW 向 (c) 波形

Fig. 8 Waveforms of the Mojiang M5.9 earthquake in UD (a), NS (b) and EW (c) directions recorded at the Jinggu station

常必要,利用大数据进行存储和应用时,宜采用文件系统和键值系统进行存储。数据应用时应用系统和软件需将 SEED 数据解压内容嵌套进去。

(3) miniSEED 数据转换成 SEED 数据格式时,将各方向的数据串连起来,增加头文件即可。

### 参考文献:

- 何加勇. 2009. 地震波形归档格式分析和转换[J]. 震灾防御技术, 4 (4): 461 - 465.
- 和跃时. 2007. CDSN 台 SEED 数据格式的软件开发[J]. 防灾减灾学报, 23 (4): 59 - 64.

- 李铂,颜启,石玉燕,等. 2009. 采用面向对象编程方法实现 SEED 地震波形格式转化[J]. 华北地震科学, 27(3): 27-30.
- 万永革,李鸿吉,姚立平,等. 1995. SEED 格式台站卷及带有可变头段节数据记录的实例剖析[J]. 地震地磁观测与研究, 16(5): 6-12.
- 王洪体,陈阳,庄灿涛. 2004. SEED 格式 STEIM2 数据压缩算法在实时地震数据传输中的应用[J]. 地震地磁观测与研究, 25(4): 14-19.
- 王晓磊,李刚,卞真付,等. 2016. 测震 NIMIED 格式数据在线实时展示方法研究和 Java 实现[J]. 震灾防御技术, 11(4): 823-829.
- 姚立平. 1993. 地震数据存储交换中的压缩技术[J]. 地震地磁观测与研究, 14(5): 35-40.
- 朱文林,张德存. 1994a. 新一代系统 CDSN 应用地震数据的格式及其使用方法[J]. 地震地磁观测与研究, 15(6): 63-69.
- 朱文林,张昱,曹永键,等. 1994b. 新一代 CDSN 台站磁带数据记录格式[J]. 地震地磁观测与研究, 15(3): 61-68.

## Compression Technology and Data Decompression of Seismic Exchange Standard Data

YANG Zhousheng, YANG Jingqiong, YAO Yuan

(Yunnan Earthquake Agency, Kunming 650224, Yunnan, China)

### Abstract

In this paper, the SEED and miniSEED format standards for seismic data storage and exchange are studied. The compression methods and coding methods of Steim1 and Steim2 in the storage format are deeply analyzed, and the differences between SEED formats at domestic and abroad are compared. Decompression and application of compressed data in SEED and miniSEED format are carried out based on the MATLAB software.

**Keywords:** SEED; miniSEED; Steim1; Steim2; data compression; encoding style; data decompression